PATENT ABSTRACTS OF JAPAN

(11)Publication number : 05-216618
(43)Date of publication of application : 27.08.1993

(51)Int.CI. G06F 3/16
G06F 3/16
G06F 15/20
G10L 3/00
G10L 3/00

(21)Application number : 04-309093
(22)Date of filing : 18.11.1992

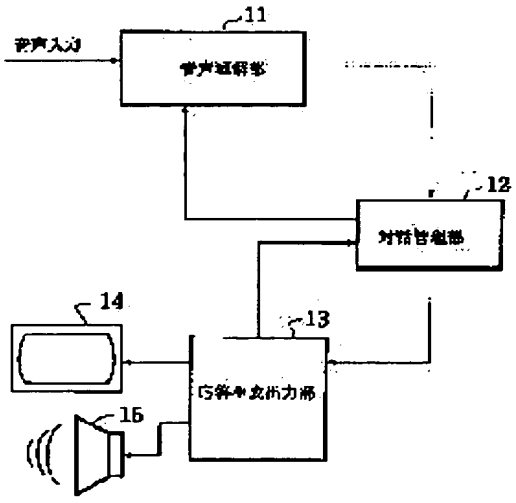
(71)Applicant : TOSHIBA CORP
TOSHIBA SOFTWARE ENG KK

(72)Inventor : TAKEBAYASHI YOICHI
TSUBOI HIROYUKI
SADAMOTO YOICHI
YAMASHITA YASUKI
NAGATA HITOSHI
SETO SHIGENOBU
SHINCHI HIDEAKI
HASHIMOTO HIDEKI

(30)Priority
Priority number : 03329475 Priority date : 18.11.1991 Priority country : JP

(54) VOICE INTERACTIVE SYSTEM

(57)Abstract:
PURPOSE: To recognize the for warding condition of the interaction and the reliability of the voice recognition by the movement of the mouth and the facial expression of the persons on a screen by displaying the personal image corresponding to an utterer from a system side.
CONSTITUTION: A voice recognition section 11 understands the inputted voice uttered by a user to extract the content of the meaning. The input meaning expression representing the understood contents is generated to be sent to an interactive management section 12. Further, an answering generation output section 13 outputs the answer generated based on the answering content information inputted from the interactive management section 12 from a speaker 15 and visually displays the person performing the voice response with the movement and expression decided based on the personal image information and answering sentences. The content visualizing information generated based on the visual information being the visual information to make the content of the conversation easy to understand for the system is visually displayed on a display 14.



LEGAL STATUS

[Date of request for examination]	21.04.1998
[Date of sending the examiner's decision of rejection]	04.09.2001
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]	
[Date of final disposal for application]	
[Patent number]	
[Date of registration]	
[Number of appeal against examiner's decision of rejection]	
[Date of requesting appeal against examiner's decision of rejection]	
[Date of extinction of right]	

(19)日本国特許庁（ＪＰ）　(12)**公開特許公報（Ａ）**　(11)特許出願公開番号

特開平5－216618

(43)公開日　平成５年(1993)８月27日

(51)Int.Cl.[5]	識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ０６Ｆ　3/16	３３０　Ｃ	8323－5B		
	Ｋ	8323－5B		
	３２０　Ｈ	8323－5B		
15/20	５０３	6798－5L		
Ｇ１０Ｌ　3/00	Ｒ	8946－5H		

審査請求　未請求　請求項の数４(全 50 頁)　最終頁に続く

(21)出願番号　特願平4－309093

(22)出願日　平成４年(1992)11月18日

(31)優先権主張番号　特願平3－329475

(32)優先日　平３(1991)11月18日

(33)優先権主張国　日本（ＪＰ）

(71)出願人　000003078
株式会社東芝
神奈川県川崎市幸区堀川町72番地

(71)出願人　000221133
東芝ソフトウェアエンジニアリング株式会社
東京都青梅市新町1385番地

(72)発明者　竹林　洋一
神奈川県川崎市幸区小向東芝町１　株式会社東芝研究開発センター内

(74)代理人　弁理士　三好　秀和　（外１名）

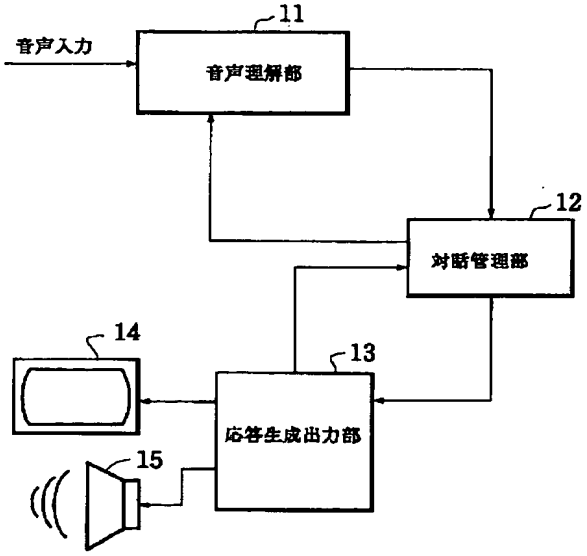
最終頁に続く

(54)【発明の名称】　音声対話システム

(57)【要約】

【目的】　本発明は、ユーザとシステムの音声対話を行う際に、システム側からユーザへの応答出力として音声応答に画面表示を併用するようにしている。

【構成】　音声入力が与えられると、入力音声の意味内容を音声理解部１１で理解し、この理解の結果に基づいて対話管理部１２により応答内容の意味的な決定を行い、この決定された応答内容に基づいて応答生成部１３により音声応答出力および画面表示出力を生成し、これら音声応答出力および画面表示出力をディスプレイ１４およびスピーカ１５より出力するように構成している。



1

【特許請求の範囲】

【請求項１】　音声入力が与えられ該入力される音声の意味内容を理解する音声理解手段と、
この音声理解手段での理解結果に基づいて対応内容の意味的な決定を行う対話管理手段と、
この対話管理手段で決定された応答内容に基づいて音声応答出力および画面表示出力を生成する応答生成手段と、
この応答生成手段で生成された音声応答出力および画面表示出力を出力する出力手段とを具備することを特徴とする音声対話システム。

【請求項２】　音声入力が与えられ該入力される音声の意味内容を理解する音声理解ステップと、
この音声理解ステップでの理解結果に基づいて応答内容の意味的な決定を行う対話管理ステップと、
この対話管理ステップで決定された応答内容に基づいて音声応答出力および画面表示出力を生成する応答生成ステップと、
この応答生成ステップで生成された音声応答出力および画面表示出力を出力する出力ステップとをから成ることを特徴とする音声対話方法。

【請求項３】　音声入力が与えられ該入力される音声の意味内容を理解する音声理解手段と、
この音声理解手段での理解結果に基づいてシステム応答出力を出力する応答出力手段と、
システムとユーザとの対話を、前記音声理解手段に音声入力が与えられるユーザ状態と前記応答出力手段からシステム応答出力が出力されるシステム状態との間の状態遷移を制御することにより、管理する対話管理手段とを具備することを特徴とする音声対話システム。

【請求項４】　音声入力が与えられ該入力される音声の意味内容を該音声入力中のキーワードを検出することにより理解する音声理解手段と、
システムとユーザとの対話の状態に応じて、前記音声理解手段により検出する音声入力中のキーワードを予め制限しておく対話管理手段と、
前記音声理解手段での理解結果に基づいてシステム応答出力を出力する応答出力手段とを具備することを特徴とする音声対話システム。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は、音声認識や音声合成を含む各種の入出力手段を利用する音声対話システムに関するものである。

【０００２】

【従来の技術】近年、文字、音声、図形、映像などのマルチメディアを入力、出力および加工処理することで、人間とコンピュータとの対話（Human-Computer Interaction）を様々な形態で行うことが可能になっている。

【０００３】特に、最近になってメモリ容量や計算機のパワーが飛躍的に向上したことでマルチメディアを扱えるワークステーションやパーソナルコンピュータが開発され、種々のアプリケーションが開発されてきているが、これらはいずれも単に種々のメディアを出し入れするだけのもので各種メディアを有機的に融合するまでに至っていない。

2

【０００４】一方、従来からの数値データに代わって文字を含む言語データが一般的になり、白黒のイメージデータはカラー化や図形、アニメーション、三次元グラフィックス、さらには動画が扱えるように拡張されてきている。また、音声やオーディオ信号についても、単なる音声の信号レベルの入出力の他に音声認識や音声合成の機能が研究開発されつつあるが、ヒューマンインターフェースとして使用するには性能が不安定で実用化は限定された分野に限られているのが現状である。

【０００５】すなわち、上述したように文字、テキスト、音声、グラフィックデータなどについては、従来の入出力処理（記録－再生）から各種メディアの理解や生成機能へと発展が続いている。換言すると、各メディアの表層的処理からメディアの内容や構造、意味的内容を扱い、人間と計算機の対話をより自然に快適に行うことを目的とした音声やグラフィックスなどのメディアの理解や生成を利用する対話システムの構築が検討されつつある。

【０００６】しかして、音声認識については、孤立単語認識から連続単語認識、連続音声認識へと発展しており、実用化のために応用を限定した方向（task-oriented）でも開発が進められている。このような応用場面では、音声対話システムとしては、音声の文字面の認識よりも音声の発話内容の理解が重要であり、例えば、キーワードスポッティングをベースに応用分野の知識を利用した音声理解システムも研究されてきている。一方、音声合成についても従来の文－音声変換（text-to-speech）システムからイントネーションを重視した対話用の音声合成システムの研究が例えば本発明者等によって行われてきており、音声対話への応用が期待されている。

【０００７】しかし、音声などのメディアの理解と生成は単なるデータの入出力と異なり、メディアの変換の際には情報の欠落やエラーが不可避である。すなわち、音声理解は情報量の多い音声パターンデータから音声の発話の内容や発話者の意図を抽出する処理であり、情報の圧縮を行う過程で音声認識エラーや曖昧性が生じる。従って、音声対話システムとしては上述した認識エラーや曖昧性などの音声認識の不完全さに対処するためシステム側からユーザに適切な質問や確認を行い対話制御によりスムーズに対話を進行する必要がある。

【０００８】ところで、対話システム側からユーザに何等かの対話をする場合、音声認識の不完全さをカバーし、計算機の状況を適確に伝えることが、使い勝手のよいヒューマンインターフェースとして重要である。とこ

ろが、従来の音声対話システムでは、音声応答として単に文を棒読みするテキスト合成が行われることが多かったためメリハリがなく聞ずらかったり、冗長であることがあった。あるいは、音声応答がなく、計算機からの応答はすべてテキストとして画面上に応答文を表示したり、あるいは図形データや映像、アイコンや数値を表示するシステムが一般的であり、視覚への負担が重くなっていた。

【０００９】このように最近では、上述したいろいろな対話システムが開発されてきているが、音声認識の不完全さに対処するためのシステム側からの応答における種々のメディアの利用に関する検討は、これまで十分になされておらず、音声認識技術の大きな問題となっていた。言い換えると、音声認識は、不安定であり、雑音や不要語に対して弱く、ユーザの意図が音声で効率よく伝えることが困難であるため、電話などの音声メディアだけにしか使えないような制約の強い場面に応用が限られていた。

【００１０】

【発明が解決しようとする課題】このように従来の音声認識、合成技術を利用した音声対話システムでは、それぞれ別個に開発された音声認識、音声合成、画面表示の各技術を単に組み合わせただけのものであり、音声の対話という観点からの十分な考慮がなされていない。すなわち、音声認識機能には、認識誤りや曖昧性があり、音声合成機能は人間の発声よりも明瞭度が悪く、イントネーションの制御も不十分のため意図や感情の伝達能力が不足しており、自然性に欠けるという根本的な問題がある。また、システム側での音声認識結果を用いて妥当な応答を生成するのも、現状の技術では不十分である。一方、応答を音声と組み合わせて画像表示することにより伝達能力が向上することが期待できるが、瞬間的に連続で時系列的な音声応答に対して二次元平面的、三次元空間的な画面表示をどのように活用し、両者のタイミングを制御するかは未解決の問題である。また、他のメディアを利用する音声対話システムとして何を表示すべきか大切な課題である。

【００１１】本発明は、上記事情に鑑みてなされたもので、システムとユーザの音声対話を効率よく、しかも正確に行うことができ、使い勝手の著しい改善を可能にした音声対話システムを提供することを目的とする。

【００１２】

【課題を解決するための手段】本発明は、音声入力が与えられ該入力される音声の意味内容を理解する音声理解手段、音声理解手段での理解結果に基づいて応答内容の意味的な決定を行う対話管理手段、対話管理手段で決定された応答内容に基づいて音声応答出力および画面表示出力を生成する応答生成手段、応答生成手段で生成された音声応答出力および画面表示出力を出力する出力手段により構成されている。

【００１３】対話管理手段は音声理解手段の理解結果に基づいて音声応答を行う発話者の人物像に関する人物像情報、音声応答に対応する発声文の応答内容テキスト情報および音声応答の内容に関連した理解内容を可視化する可視化情報をそれぞれ応答内容として出力するようにしている。

【００１４】応答生成手段は対話管理手段より出力される音声応答を行う発話者の人物像情報に基づいて人物像の動作および表情の少なくとも一方の画面表示出力を生成するようにしている。

【００１５】また、応答生成手段は対話管理手段より出力される音声応答を行う発話者の人物像情報に基づいて人物像の動作および表情の少なくとも一方の画面表示出力を生成するとともに各画面表示に対応する音声の感情または強弱を有する音声応答出力を生成するようにしている。

【００１６】さらに、人の動きに関する人状態を検出する人状態検出手段を有し、該人状態検出手段の検出結果に基づいて対話管理手段にて応答内容の意味的な決定を行うようにしている。

【００１７】そして、音声入力が可能か否かのアイコンを表示可能にしている。

【００１８】また、本発明は、音声入力が与えられ該入力される音声の意味内容を理解する音声理解手段と、この音声理解手段での理解結果に基づいてシステム応答出力を出力する応答出力手段と、システムとユーザとの対話を、前記音声理解手段に音声入力が与えられるユーザ状態と前記応答出力手段からシステム応答出力が出力されるシステム状態との間の状態遷移を制御することにより、管理する対話管理手段とを具備することを特徴とする。

【００１９】さらに本発明は、音声入力が与えられ該入力される音声の意味内容を該音声入力中のキーワードを検出することにより理解する音声理解手段と、システムとユーザとの対話の状態に応じて、前記音声理解手段により検出する音声入力中のキーワードを予め制限しておく対話管理手段と、前記音声理解手段での理解結果に基づいてシステム応答出力を出力する応答出力手段とを具備することを特徴とする。

【００２０】

【作用】この結果、本発明は、ユーザとシステムとの間の対話を行う際に、音声認識、音声応答に加えて、システム側からユーザへの応答出力として応答の画面表示を併用するようになる。この時、システム側からの音声の発話者に対応する人物像の表示を行うことにより、発話者のイメージがシステムの機能を代表するようになり、ユーザは画面上の人物に向かって発声することを自然に行うことができ、また、画面上の人物の口の動きや表情で対話の進行状況や音声認識の信頼性を把握できる。

【００２１】一方、システムからの応答内容に関して

は、応答文を表す文字列を表示するほか、対象物（例えば商品、概念などの物や事）や数などについては図形などで表示することから、応答内容をユーザに素早く伝えることもできる。さらに、音声認識は、誤認識や曖昧性が多発する不完全なものであり、音声の内容を理解する際に、ユーザの意図しない誤った情報が計算機側に伝えられることがあるが、音声応答の他に、視覚による各応答表示を並行して用いることで、音声対話の効率を大幅に向上させ、自然性や使い勝っての改善を可能にできる。また、音声合成音は、自然音声に比べて低いので、発話者の表情、応答文、応答内容の視覚化を併用することは対話の改善に極めて有用である。

【００２２】

【実施例】以下、本発明の一実施例を図面に従い説明する。

【００２３】図１は音声対話システムとしての画面表示を加えたシステムの概略構成を示している。

【００２４】音声対話システムは、入力される音声の意味内容を理解する音声理解部１１、音声理解部１１での理解結果に基づいて応答内容の意味的な決定を行う対話管理部１２、対話管理部１２で決定された応答内容に基づいて音声応答出力および画面表示出力を生成する応答生成出力部１３、応答生成出力部１３で生成された画面表示を出力する画面表示出力部１４および音声応答を出力する音声出力部１５により構成されている。

【００２５】音声理解部１１は、音声の文字面の認識、すなわち単語や文の認識でなく、ユーザの発話した入力音声の理解を行い意味内容を抽出する。そして、理解した意味内容を表す入力意味表現を生成し対話管理部１２に送る。

【００２６】対話管理部１２では入力音声の入力意味表現に対して、対話の履歴や現在の対話の状態に関する情報と対話の進行方法や応用分野の知識を用いて応答内容の意味的な決定を行ない、音声応答に対応する発声文の応答内容情報を応答生成出力部１３に出力する。

【００２７】さらに、対話管理部１２では、省略や指示代名詞を含む話し言葉を処理し、音声理解の性能向上や処理量の削減とともに自然な対話を可能にしている。また、対話管理部１２は、ディスプレイ１４に表示出力されて音声応答を行う発話者の人物像情報、および音声応答の内容に関連した理解内容の可視化する情報である可視化情報を応答生成出力部１３に出力する。

【００２８】また、対話管理部１２で生成された出力意味表現を音声理解部１１へ送り、出力意味表現から次の発話のキーワードや構文的意味的規則を絞り、次の発話の音声理解性能の向上をはかることが可能となる。

【００２９】応答生成出力部１３は対話管理部１２から入力された応答内容情報に基づいて生成された応答文を合成音声でスピーカー１５より出力するとともに、人物像情報および応答文に基づいて動作や表情が決定された音声応答を行う人物像をディスプレイ１４に視覚的に表示し、また、それまでの対話によりシステムが理解した内容を分かりやすく可視化するための情報である可視化情報に基づき生成された内容可視化情報をディスプレイ１４に視覚的に表示して、複数のメディアを利用してマルチモーダル的に応答をユーザに提示する。つまり、オーディオ情報と視覚情報を併用してユーザに提示することにより音声対話システムのヒューマンインターフェースが改善され、自然な対話が可能となる。

【００３０】また、応答生成部１３から現在応答を出力中である旨の情報を対話管理部１２に送る。対話管理部１２では、上記情報を音声理解部１１へ送り、例えば入力音声の終始端検出処理や、キーワード検出処理のタイミングを制御することにより音声理解性能の向上をはかることが可能となる。

【００３１】次に、上述した音声対話システムの各部について、ここでは応用としてファースト・フードでの注文タスクを想定してさらに詳しく説明する。

【００３２】まず音声理解部１１について説明する。音声理解部１１については、先に述べたようにここでの役割は、テキスト入力や音声ワードプロセッサのように文字面を認識するのではなく音声の意味内容や発話者の意図や状況を理解することを目的としている。

【００３３】この場合、不特定ユーザを対象とする券売機、航空機や列車の座席予約システム、銀行の現金自動引出機などでは、話者間の音声の違い、不要語、口語の話し方の違い、雑音の影響などにより実際に音声認識技術を応用しても十分な認識性能が期待できないことがあり、とくに発話された文の高精度認識に困難を極めている。これについて、連続発声された音声から、まずキーワードの候補列を解析して発話内容を理解する方法が例えば文献（坪井宏之、橋本秀樹、竹林洋一：“連続音声理解のためのキーワードラティスの解析”日本音響学会講演論文集、１－５－１１、ｐｐ．２１－２２、１９９１－１０）に提案されており、この方法を用いれば、限定した応用では、利用者の発話に極力制限を設けずに、自由な発声を高速に理解できるようになる。

【００３４】図２は、上述したキーワードを利用した音声理解部１１の概略構成を示している。

【００３５】この場合、音声理解部１１は、キーワード検出部２１と構文意味解析部２２から構成している。そして、キーワード検出部２１は、音声分析部２１ａとキーワードスポッティング処理部２１ｂにより構成し、構文意味解析部２２は文始端判定部２２ａ、文候補解析部２２ｂ、文終端判定部２２ｃ、文候補テーブル２２ｄにより構成している。

【００３６】キーワード検出部２１では、音声分析部２１ａにより入力音声をローパスフィルタ（ＬＰＦ）を通し標本化周波数１２ｋＨｚ、量子化ビット１２bitsでＡ／Ｄ変換してディジタル信号に変換し、次いで、スペク

トル分析、さらにはＦＦＴを用いたのちに周波数領域での平滑化をそれぞれ行い、さらに対数変換を行って１６チャンネルのバンドパスフィルタ（ＢＰＦ）より８msごとに音声分析結果を出力し、この出力に対してキーワードスポッティング処理が実行される。この場合のキーワードスポッティング処理は、例えば文献（金沢、坪井、竹林：”不要語を含む連続音声中からの単語検出”電子情報通信学会音声研究会資料、ｓｐ９１－２２、ｐｐ．３３－３９、１９９１－６）に開示された方式により行うことができる。

【００３７】これによりキーワード検出部２１では、連続した入力音声よりキーワードの候補系列（ラティス）を抽出するようになる。図３は、ファースト・フード店での店頭での注文のやりとりを音声対話で行うのに適用した場合の連続入力音声「ハンバーガとポテトとコーヒー３つ下さい」より抽出されたキーワードの候補系列の例を示している。

【００３８】なお、上述の音声分析やキーワード検出処理は、他の文献（“高速ＤＳＰボードを用いた音声認識システムの開発“日本音響学会講演論文集、３－５－１２、１９９１－３）にあるようなＤＳＰボードを用いることでリアルタイム処理も可能である。

【００３９】次に、このようにして検出されたキーワード候補系列が構文意味解析部２２により構文意味解析され、図４に示すような音声入力に対する入力意味表現が求められる。

【００４０】ここでは応用をファースト・フードのタスクに限定しており、フレーム形式の入力意味表現は、入力発話が注文処理の種類を表すＡＣＴフレームと注文内容を表す注文品フレームから構成される。そして、ＡＣＴフレームには“注文”、“追加”、“削除”、“置換”など、注文に関する処理についての意味情報が表現され、一方、注文品フレームには、品名、サイズ、個数のスロットからなる、注文品の内容を表現できるようにしている。

【００４１】即ち、キーワード検出部２１で得られキーワードラティスは構文意味解析部２２に送られる。構文意味解析部２２は、文始端判定部２２ａ、文候補処理部２２ｂ、文終端判定部２２ｃから成り、文候補テーブル２２ｄを持つ。構文意味解析部２２は、キーワードラティス中の個々の単語を左から右に向かって処理していく。

【００４２】文始端判定部２２ａは、現在処理している単語が文の始端となりうるか否かを構文的意味的制約により判定する。もしそれが文の始端となりうるならば、その単語を新しい部分文候補として、文候補テーブル２２ｄに登録する。

【００４３】文候補解析部２２ｂは、当該単語および文候補テーブル２２ｄ中の各部分文候補に関して、構文的意味的時間的制約から、それらが接続しうるか否かを判定する。もし接続しうるならば、部分文候補をコピーし、それに入力単語を接続し、それを文候補テーブル２２ｄに登録する。

【００４４】文終端判定部２２ｃは、直前に文候補解析部２２ｂで処理された部分文候補が、構文的意味的に文として成立しうるか否かを判定し、成立するならばその部分文候補のコピーを構文意味解析部２２の出力として出力する。

【００４５】出力された文候補は、構文解析と同時に意味解析が行われており、従ってこれがそのまま入力意味表現を意味する。以上の処理は入力に対してバイブライン的に行われる。かくして、入力音声に対する複数の入力意味表現を得ることができる。

【００４６】この場合のファーストフード・タスクでは、キーワードとして図５に示すようなものを用いているが、対話の状況によっては、別の発話が同じ意味となることもある。すなわち、キーワードに基づく音声理解では、“１つ”と“１個”は同じ意味表現であり、また“下さい”を“お願いします”も同じ意味表現になることがあり、表層的な文字面の入力音声の表現とは異なってくる。ここが音声認識と音声理解の相違点であり、本発明で扱う音声対話システムにおいては、応用分野の知識を用いたタスク依存の音声理解処理が必要となっている。

【００４７】次に対話管理部１２について説明する。本発明による音声対話システムでは、図１に示すように音声理解部１１から出力される入力意味表現は対話管理部１２に送られ、対話の知識や応用分野の知識さらに対話の履歴や状態の情報を用いて応答内容の意味的な決定を行ない、確認応答のための応答内容情報てある出力意味表現を生成し応答生成出力部１３に出力する。なお、出力意味表現は図６に示すように、入力意味表現と同様にフレーム形式の表現を用いている。

【００４８】本実施例では、入力音声の一つの発話の内容表現として図４に示す入力意味表現を用いているが、さらに、対話開始からのシステムが理解した内容として、それまでの注文の内容を記憶する注文テーブルを図７に示すように別途用意している。また、対話の履情として対話進行にともなう注文テーブルの変化を図８の例に示すような一つ前の質問応答時点の注文テーブル（旧注文テーブル）の形で用意している。さらに、対話の状況を表す対話状況情報を対話管理部１２に保持している。こ対話状況情報は、現在の対話の状態、次に遷移する状態、対話の繰返し回数、確信度、強調項目、対話の履歴等の情報を含むもので、後述する応答生成出力部１３において人物像情報として利用されるものである。

【００４９】注文テーブルは、入力意味表現のＡＣＴ情報と注文内容に基づき書き替えられたもので、形式は入力意味表現からＡＣＴ情報を取り去った注文内容のみのテーブルである。つまり、この注文テーブルは、対話を

開始してからそれまでの対話で理解した内容を反映したものである。また、旧注文テーブルは注文テーブルと同一の構成であり、一つ前の対話時点での質問応答での注文テーブルを保持し、対話の履歴情報として注文テーブルの状態を記録するものである。

【００５０】このように対話管理部１２では、入力音声の意味表現（入力意味表現）および対話の履歴情報（旧注文テーブル）、対話システムの状態に基づき、対話の進行方法や応用分野の知識を用いて応答出力の内容を表す応答内容情報（出力意味表現）を生成する。つまり、入力意味表現と注文テーブルを参照し、その時点のシステムの状態（ステート番号）に依存した処理を行い、応答生成の内容と応答ＡＣＴからなる応答生成の内容を表現した出力意味表現を生成するようにしている。上述したようにこの場合の出力意味表現は、入力意味表現と同様にフレーム形式の表現を用いている。さらに、対話の履歴情報（旧注文テーブル）と対話システムの状態に基づき、対話状況情報を生成し、応答生成出力部１３が応答画面表示の人物像情報として参照できるようにしている。

【００５１】図９は、対話管理部の内部における状態遷移の一例を示している。

【００５２】この例では対話の進行方法や応用分野の知識に基づいた状態遷移の表現により対話を管理進行するようにしている。対話管理部１２は、大きくユーザ７２とシステム７１のそれぞれの状態に二分される。

【００５３】ここで、ユーザ７２の状態の役割は、ユーザの発話の入力意味表現に応じてシステム７１の状態に遷移することであり、一方、システム７１の状態の役割は、理解した発話内容に応じて、注文テーブルの内容を変更し、応答の出力意味表現を出力して、対話の流れを進行し、ユーザ７２の状態に遷移することである。このように、システムの内部状態を二分して持つことにより、ユーザとシステムとの多様なやり取りが表現でき、柔軟な対話の進行が可能となる。

【００５４】又、この時用いられる対話状況情報は、処理中の対話管理の状態名と部分的な対話の繰り返し回数を表し、進行中の対話の状態名、次に遷移する状態名、同一の話題について同じ質問を繰り返すなどの部分的な対話が繰り返される回数が逐次記録され容易に参照できるようになっており、システムの状態を自然に分かりやすく伝えるために、応答生成出力部１３の人物像の表情、動き、および音声応答の感情、強調などの人物像情報として利用し応答生成出力に利用される。

【００５５】さて、図９では、対話管理部１２において、ユーザ（客）の存在が検知されると、システム７１の初期状態Ｓ０から対話がスタートして、挨拶、注文要求に関する出力意味表現を生成し、応答生成出力部１３に送られユーザ７２の初期状態Ｕ０に遷移する。さらに、対話の履歴情報（旧注文テーブル）は初期化され、システムの状態の状態Ｓ０から状態Ｕ０への遷移に基づき、対話状況情報が生成される。この対話情況情報は、応答生成出力部１３が応答画面表示の人物像情報として参照できるようにしている。

【００５６】すると、応答生成部１３では、この出力意味表現に基づいて、システム状態、対話の履歴情報、注文テーブルを参照しながら音声応答、人物像、テキスト、内容可視化情報を生成する。

【００５７】この時、ユーザ７２の初期状態Ｕ０では、次の発話の入力意味表現ＡＣＴ情報が“注文”であるときには、一般的な注文の流れでシステム７１の対話進行状態ＳＰとユーザの対話進行状態ＵＰの間の遷移へと移行する。

【００５８】一方、入力意味表現のＡＣＴ情報が、注文以外の場合には、そのユーザの発話は予期していないユーザの発話と見なされ、システム７１の対話修正状態Ｓ１０に遷移する。

【００５９】もし、システム７１の対話修正状態Ｓ１０に遷移した場合には、システム７１は入力意味表現、注文テーブルやその履歴情報を用いて、ユーザ７２から受け取った音声入力が予期せぬ内容であったり、良く聞こえなかった旨を状況に応じて適当な応答でユーザ７２に伝えたり、注文の内容を一品目づつ詳細に確認するための出力意味表現を出力し、ユーザ７２の対話進行状態ＵＰに遷移するようになる。すると、応答生成部１３では、この出力意味表現に基づいてシステム状態、対話の履歴情報、注文テーブルを参照しながら音声応答、人物像、テキスト、内容可視化情報を生成する。

【００６０】このようにしてシステム７１とユーザ７２の間でやり取りが続き、ユーザ７２での発話とシステム７１での応答が行われ状態が遷移して行くが、ユーザ７２が目的とする注文を終えた場合、すなわちシステム７１の全注文の確認応答に対して、ユーザ７２が肯定を意味する発話をした場合、システム７１の終了状態Ｓ９に移り、対話を終了する。

【００６１】図１０はユーザの状態の処理のフローチャートを示している。

【００６２】この場合、最初に複数の入力意味表現を読み込み（ステップＳ８１）、省略表現の推論を行い（ステップＳ８２）、各入力意味表現の確からしさに関する得点付け（スコアリング）を行う（ステップＳ８３）。次いで、上記入力意味表現候補の中から最もスコアの高い入力意味表現を選択し（ステップＳ８４）、発話アクトを決定し（ステップＳ８５）、発話アクトに基づきシステムの状態に遷移するようになる（ステップＳ８６）。

【００６３】一方、図１１はシステムの状態の処理のフローチャートを示している。

【００６４】この場合、最初に入力意味表現に基づき注文テーブルの内容を変更し（ステップＳ９１）、その時

点におけるシステムの状態を考慮して出力意味表現を生成し（ステップＳ９２）、応答内容を出力し（ステップＳ９３）、ユーザの状態へ遷移するようになる（ステップＳ９４）。応答生成部１３では、生成された出力意味表現に基づいて、音声応答、人物像、テキスト、内容可視化情報を生成する。

【００６５】以上のように、本実施例システムにおいては、ユーザとシステムとに分けて、相手側からメッセージを受け取った場合に、種々の知識、状況、メッセージ内容を考慮した処理が可能であり、柔軟で尤もらしい処理が可能なる。

【００６６】次に、図１２は、本実施例システムにおける対話処理の具体例を示すものである。

【００６７】この場合、システムでは、ユーザの発話に対し前回の状態の図１２（ｂ）に示す注文テーブルと図１２（ａ）に示す出力意味表現が提示されているとすると、これらに基づいて、図１２（ｃ）に示すように「御注文はハンバーガ１つ、コーヒーを２つ、コーラの大を４つですね」の確認メッセージを生成し、これに基づく音声応答、確認のテキスト、注文テーブルの品物の絵と個数を、音声メディアと視覚メディアを用いてマルチモーダル的にユーザに提示するようになる。

【００６８】これに対して、ユーザが図１２（ｃ）に示すように「コーラを１つ追加して下さい。」と音声入力したとすると、図１に示す音声理解部１１では、上述したようにキーワード候補の検出を行った後に、キーワード候補系列（キーワードラティス）の解析（パージング）を行い、ユーザの発話に対する図１２（ｄ）と図１２（ｅ）に示す入力意味表現候補１と入力意味表現候補２を得る。

【００６９】ここでの各候補は、確からしさについてのスコア（得点）Ｄを持っており、入力意味表現候補１はＤ１、入力意味表現候補２はＤ２のスコアを有し、それぞれスコア順に並べられる。

【００７０】この場合は、スコアＤ１の入力意味表現候補１では、ＡＣＴ情報は追加、品名はコーラ、サイズは不定、個数は１となっており、スコアＤ２の入力意味表現候補２では、ＡＣＴ情報は削除、品名はポテト、サイズは大、個数は２となっている。

【００７１】そして、ユーザの状態での処理は図１３に示すように実行される。

【００７２】まず、入力意味表現候補１については、前回の出力意味表現のコーラのサイズが大であったことを参照し、コーラの今回の追加注文もサイズが大であると推論することで省略表現を補う（ステップＳ１１１）。入力意味表現候補２については、とくに省略はないのでこの推論は行われない（ステップＳ１１３）。

【００７３】次に、妥当性のチェックを行う。すなわち、注文テーブルの内容と入力意味表現候補を照らし合わせ矛盾の有無を調べる（ステップＳ１１２、ステップ１１４）。

【００７４】この例では、入力意味表現候補２については、入力意味表現の発話ＡＣＴが“削除”で品名がポテトの大にもかかわらず、注文テーブルにポテトはないのでスコアＤ２がＤ２′＝Ｄ２×α（α＜１．０）の処理により、小さくされる処理を行う。

【００７５】次に、入力意味表現候補のスコアを比較する（ステップＳ１１５）。この場合、Ｄ２′＜Ｄ１′なので候補１を選択する。一方、ユーザからの入力の発話ＡＣＴは“追加”であると決定し（ステップＳ１１６）、追加確認を行うシステムの状態ＳＡに遷移するようになる（ステップＳ１１７）。ここで、追加確認を行うシステムの状態ＳＡは、注文テーブルを入力意味表現に基づいてコーラの大を１つ追加するように書き換える。

【００７６】この場合、システムでの状態の処理は図１４に示すように実行される。

【００７７】即ち、この状態でシステム側では、ユーザへの確認応答を図１５（ｂ）に示す応答ＡＣＴリストから選択し出力意味表現を決定する。この例では、入力意味表現のＡＣＴが追加であるので応答ＡＣＴは第４番目の追加確認が選択され、これらの情報を用いて応答出力（応答文）が決定され出力が行われる。

【００７８】まず、注文テーブルにコーラの大を１つ追加する（ステップＳ１２１）。そして、ここでの図１５（ａ）に示す入力意味表現の発話ＡＣＴは追加なので、応答ＡＣＴを追加確認として選択し、これら情報から図１５（ｃ）に示す出力意味表現が求められる（ステップＳ１２２）。次いで、応答出力（応答文）を決定して出力する（ステップＳ１２３）。この場合の応答文は、図１６に示すシステムの応答の表現例に基づいて決定され、例えば「確認します。コーラの大を１つ追加ですね。」のように出力される。そして、ステップＳ１２４に進み、追加確認の応答ＡＣＴを受けたユーザの状態ＵＡに遷移し、ユーザの状態での処理が行われるようになる。

【００７９】尚、対話管理部１２は、上述のように求めた出力意味表現と共に、音声理解部１１から受け取ったユーザの発声速度と各キーワードの尤度を応答生成出力部１３に送る。

【００８０】ここで、ユーザの発声速度は次のように求められる。即ち、図２における音声理解部１１のキーワードスポッティング処理部２１ｂで、得られたキーワードとその始終端、およびそれらを基に得られるユーザの発声速度を検出する。また、音声理解結果の各キーワードの尤度を入力意味表現とともに対話管理部１２に出力する。ユーザの発声速度は、キーワードスポッティング処理部２１ｂから得られる単語の始終端とその単語が分かれば、例えば図１７のように求めることができる。即ち、ユーザの発声から３つのキーワード“ハンバー

ガ”、“ポテト”、“ください”がそれぞれ始端ｔ１かつ終端ｔ２、始端ｔ３かつ終端ｔ４、始端ｔ５かつ終端ｔ６と検出されたとき、これらキーワードのモーラ数は＊６，３，４であることから、ユーザの平均発声速度は

【数１】

$$\{6/(t2-t1)+3/(t4-t3)+4/(t6-t5)\}/3$$

のように算出できる。

【００８１】このようにして得られたユーザの発声速度と尤度は、入力意味表現と共に対話管理部１２へ入力される。

【００８２】対話管理部１２は図９から図１６で説明された処理に基づき生成された出力意味表現に、音声理解部１１から入力されたユーザの発声速度と、キーワードの尤度を応答生成出力部１３に入力する。このときの出力意味表現の例を図１８に示す。

【００８３】次に応答生成出力部１３について説明する。本発明による応答生成出力部１３では応答内容情報である出力意味表現、対話情況情報と対話履歴情報からなる人物像情報、およびそれまでの対話によりシステムが理解した内容を分かりやすく可視化するための可視化情報に基づいて、音声応答、人物像、テキスト、内容可視化情報を生成出力する。ここで、音声応答、人物像、テキストは出力意味表現と人物像情報に基づいて、その対話状況を考慮して伝える内容をわかり易く呈示できるように表情や感情・強調を持って生成される。また、内容可視化情報はシステムの理解している対話の内容を表示して、対話の進行状況をわかり易くするためのものであり、対話管理部１２から出力される可視化情報に基づいて生成出力されるものである。

【００８４】図１９は応答生成出力部１３の構成の一例を示している。応答生成部１３は応答文生成部１３１、人物像表情決定部１３２、人物像生成部１３３、感情・強調決定部１３４、音声応答生成部１３５、内容可視化情報生成部１３６、出力統合制御部１３７から構成される。

【００８５】応答生成出力部１３は対話管理部１２から出力意味表現と人物像情報を受けとり、応答文生成部１３１で応答する文とその構造情報を生成する。生成された応答文と対話管理部１２からの人物像情報に基づき、人物像表情決定部１３２では画面表示で音声応答する人物像の表情を決定し、決定された表情の人物像を人物像生成部１３３で生成し出力統合制御部１３７に出力する。また、生成された応答文と文構造情報および対話管理部１２からの人物像情報に基づき、音声応答の感情表現や強調する部分を感情・強調決定部１３４で決定し、感情や強調を持つ音声応答を音声応答生成部１３５で生成し出力統合制御部１３７に出力する。さらに、生成された応答文はテキスト情報として出力統合制御部１３７に出力する。また、応答内容に関連した理解内容を可視化して表示するために、応答生成出力部１３は対話管理部１２から出力される可視化情報を受けとり、内容可視化情報生成部１３６で内容可視化情報を生成し出力統合制御部１３７に出力する。

【００８６】出力統合制御部１３７は表情を持つ人物像、感情や強調を持つ音声応答、テキスト情報、内容可視化情報を各部から受けとり、時間的な呈示順序を制御しながら、画面表示出力部１４と音声出力部１５に出力して利用者に応答内容を統合して表示する。

【００８７】次に、応答生成出力部１３の各部の動作を図１９に基づき説明する。

【００８８】まず、応答生成出力部１３の各部で処理される情報について説明する。

【００８９】出力意味表現は図６に示したような入力意味表現と同様なフレーム形式であり、ＡＣＴ情報は応答におけるアクションを示している。

【００９０】人物像情報は画面表示出力部１４に表示される音声応答する人物像の表情や音声応答の感情・強調の情報であり、図２０に示すような構造である。システム状態番号、ユーザ状態番号は対話管理部１２の対話処理においてシステム状態からユーザ状態へ遷移して出力意味表現を生成する際のそれぞれの状態の番号を示している。図２０に示すＳＰ１、ＵＰ３はそれぞれ図１９の対話状態遷移のシステム側の状態集合ＳＰの１つ状態を、ユーザ側の状態集合ＳＵの１つの状態を示している。繰り返し回数は対話の中で部分的に同じ質問を繰り返して行なったり、同じ内容について繰り返して確認を行なうような場合の回数である。強調項目は出力意味表現の中で特に確認する必要がある場合の項目を示す。確信度は出力意味表現に基づいて対応する内容の確信度を示し、対話管理部１２で入力意味表現の尤度に基づてい入力意味表現の解釈を行なった際に得られるスコアＤである。応答文生成部１３１は、対話管理部１２で生成された出力意味表現から応答文とその文構造を生成する。文生成には、一般に書き換え規則を使うもの、穴埋めによるもの、木構造から合成する方法、意味構造から合成する方法が知られているが、ここでは穴埋めによる方法で説明する。

【００９１】出力応答文の生成は、例えば図２１のようにＡＣＴ情報ごとに品目、サイズ、個数を埋め込む穴の空いた応答文型とその文構造を用意しておき、図２２（ａ）に示すフローチャートにしたがって出力意味表現をもとに空きを埋める方法で実現できる。すなわち、まずステップＳ１４１で繰り返しの回数を示す変数ｎを０に設定し、次にステップＳ１４２で出力意味表現の品目数を変数Ｍにセットする。図２２（ｂ）の出力意味表現の場合には、Ｍは２である。次に、ステップ１４３で注文一品目分の品名、サイズ、個数を応答文に埋め込む。次にステップＳ１４４で繰り返し変数ｎを加算しなが

ら、ステップＳ１４５により埋め込みが完了するまで繰り返す。図２２（ｂ）の出力意味表現を図２２（ｃ）の応答文型に埋め込むと、図２２（ｄ）のように「確認します。コーラの大は１つ、ポテトの小は３つですね。」と応答文が得られる。

【００９２】人物像表情決定部１３２は、応答文生成部１３１で生成された文と対話管理部１２から入力される人物像情報から人物像の表情を決定する。人物像表情決定部１３２の一例を図２３に示す。システム状態番号、ユーザ状態番号、繰り返し回数、確信度は人物像情報から得られるものであり、あらかじめそれぞれの値に対しての人物像とその表情をテーブルの形式で表現したものである。例えば、繰り返し回数が一回までの場合の確認で確信度が高い場合には普通の表情で確認を行い、確信度が低い場合には戸惑ったような表情で確認を行なうようになっている。

【００９３】人物像生成部１３３は、人物像表情決定部１３２から出力された人物像と表情の情報から画面に表示する画像を生成する。この時、表示時間や人物像を変化させるための制御が行なわれる。例えば、人物像が音声応答する際の口の動作や挨拶する時のおじぎの動作が生成できるように、静止画を用いる場合には複数の画像が用意され、動画を用いる場合には連続した動作の人物像と表情の動画が指定される。また、コンピュータグラフィックスを利用する際には指定された動作の画像が生成される。

【００９４】感情・強調決定部１３４は、人物像情報から応答する音声の強調や感情を決定する。感情・強調決定部１３４の一例を図２４に示す。人物像表情決定部１３２と同様の表現形式であり、システム状態番号、ユーザ状態番号、繰り返し回数、確信度から、あらかじめそれぞれの値に対しての人物像と音声応答の感情をテーブルの形式で表現したものである。例えば、繰り返し回数が一回までの場合の確認で確信度が高い場合には普通の音声で確認を行ない、確信度が低い場合には戸惑ったような音声で確認を行なうようになっている。また、確認する場合に強調して利用者に確認内容をわかり易く伝えるために人物像情報には強調項目がある。これは対話管理部１２で応答内容を出力意味表現として生成する際に確認すべき項目を決定したものである。感情・強調決定部１３４では応答文中の強調すべき項目を出力意味表現からとりだして次の音声応答生成部１３５に伝える。

【００９５】音声応答生成部１３５は、応答文生成部１３１と感情・強調決定部１３４からの出力に基づき音声合成を行なう。音声の生成方式としては従来からある録音編集型なども利用可能であるが、本実施例では強調や感情を持つ応答に特徴があり、音声生成部の制御により実現するためには音声規則合成が望ましい。

【００９６】音声応答生成部１３５の構成の一例を図２５に示す。音声応答生成部１３５は、音韻処理部１５１、韻律処理部１５２、制御パラメータ生成部１５３、音声波形生成部１５４からなる。

【００９７】ここでは、感情・強調決定部１３４から入力される強調する語句（句）と感情の種類、および生成された応答文とその構造を基に音韻処理部１５１と韻律処理部１５２において各々音韻処理、韻律処理を行なって、音声波形生成部１５４で使用される制御パラメータの時系列を制御パラメータ生成部１５３から音声波形生成部１５４に出力する。

【００９８】音韻処理部１５１は、応答文生成部１３１で生成された応答文とその文構造を基に、鼻音化や無声化、連濁といった一般に良く知られた音韻規則に従い出力応答文の読みを決定、単音記号列を出力する。

【００９９】韻律処理部１５２では応答文とその構造、強調する語の情報および感情の種類を基に、基本周波数パターンやパワー、継続時間、ポーズの位置などの韻律成分を決定する。

【０１００】特に基本周波数パターン生成は、図２６のモデルに示すように、点線で示したあらかじめ強調しない場合と実線で示した強調した場合の各応答文のアクセント成分やフレーズ成分の多きさを分析して記憶しておき、合成時に語句、句にその成分を使うことで実現できる。また、平叙文と疑問文と命令文というように文の種類を分類し、文の種類毎にアクセントやフレーズの規則を作成してもよい。例えば文献（広瀬、藤崎、河井“連続音声合成システム－特に韻律的特徴の合成－”、日本音響学会音声研究会資料Ｓ８５－４３（１９８５））のように、単語のアクセント型、文の切れ目からの語の位置、修飾関係から平叙文の規則を決めることができる。

【０１０１】感情を伴った韻律は、文献（K.Sheahan,Y.Yamashita,Y.Takebayashi,“Synthsis of Nonverbal Expressions for Human-Computer Interaction”日本音響学会講演論文集２－４－６（１９９０．３））に述べられているように、おもに基本周波数の変化の割合とダイナミックレンジ、発声時間長、エネルギーによって制御される。従って、図２７に示すように喜びの場合には感情を伴わない場合に対してアクセントを１．２倍、発声時間を０．９倍、エネルギーを２dB大きくし、図２８に示す悲しみの場合にはアクセントを０．９倍、発声時間を１．１倍、エネルギーを２dB小さくする。これにより喜びを伴ったときは、一語一語はっきりと、やや早口になった音声が合成でき、悲しみを伴ったときは抑揚が少なく、やや遅い合成音を生成することが可能である。

【０１０２】基本周波数の制御は図４１で用いたものに限らず、直線近似を用いた方法や音の高低のレベルで基本周波数パターンを表現する方法があり、ここに述べたものに限らず、発明の主旨を逸脱しないならば種々の方法を利用してもよい。

【０１０３】制御パラメータ生成部１５３では、音韻処理部１５１と韻律処理部１５２からの音韻シンボルと韻

律シンボルを基に、音声波形生成部１５４で使う制御パラメータを決定する。この制御パラメータ生成部１５３では発声速度の制御も行なうため、ユーザの発声速度に合わせて音声を合成することが可能となり、ユーザの発声のペースで対話を進行することも可能である。

【０１０４】このため制御パラメータ生成部で得られた発話時間長は人物像の口動作と音声応答の同期をとるために出力統合制御部１３７に出力される。

【０１０５】尚、この応答生成出力部１３では、応答文の生成はすでに述べたような応答文生成部１３１、感情・強調決定部１３４、音声応答生成部１３５により行われるが、ここで、発声速度は応答文の長さに反映するために参照する。テンポの良い対話がなされている時には、応答は短い方が良く、ユーザが戸惑うなどの理由でゆっくり発声する時には、丁寧に省略などせずに応答するのが良いからである。例えば発声速度が９モーラ毎秒より速ければ、短い応答文型を選ぶようにすることで、これは実現される。

【０１０６】また、対話管理部１２から与えられる各キーワードの尤度は、例えば確認の場面で文末の「ですね／ですか」を使い分けるのに利用される。すなわち、キーワードの平均尤度が例えば設定域値０．５より低い、もしくはどれかのキーワードの尤度が非常に低い時には「ですか」を使い、尤度が高い時には「ですね」を使う。これにより、他の応答出力に加え、応答文からも計算機の理解の程度が分かるようになり、ユーザが対話を行ないやすくなる。

【０１０７】なお、「ですね／ですか」は文型のテーブルとして持たずに、文型を決定してから変更できるようにしてもよい。また、「でございます／でございますか」のように、丁寧な応答か否かの情報を使うなどして別の言葉を使用しても良い。

【０１０８】音声波形生成部１５４は、例えば図２９に示すようなホルマント型合成器による規則合成を利用する。これれは例えば、標本化周波数を１２ｋＨｚ、８ｍｓｅｃごとに合成パラメータを更新し、音源にはインパルスにローパスフィルターをかけたものを利用することで音声合成ができる。しかし、合成器の構成、音源の種類、標本化周波数等も一般的に知られものを利用することが可能である。

【０１０９】尚、この図２９に示すホルマント型合成器から成る音声波形生成部１５４においては、制御パラメータ合成器１６９から入力された制御パラメータがインパルス発生器１６１、雑音発生器１６２、ローパスフィルタ１６３Ａ，１６３Ｂ、振幅制御器１６７、ハイパスフィルタ１６８、共振器１６６Ａ，１６６Ｂにそれぞれ分配される。

【０１１０】可視化情報は、対話中にシステムに伝えた内容、システムが理解している内容、システムの状態などの情報であり、この可視化情報を基に内容可視化情報生成部１３６が内容可視化情報を生成しユーザに視覚的に呈示することにより、システムの状態や理解内容をシステムと利用者が共有することが可能となり、対話を自然にわかり易く進めることができる。

【０１１１】本実施例では注文テーブルを可視化情報として用いている。注文テーブルには既に利用者が注文したすべての品目、サイズ、個数が記録されており、対話の各時点での注文内容を確認することができる。これにより、例えば品目が多い注文を行なった時に、それぞれの品目とサイズ、個数を音声応答だけで時間的に連続して聞く場合よりも視覚的に表示して並列的に注文の内容を伝えることが可能となる。内容可視化情報生成部１３６はこの可視化内容情報に基づき画像の生成を行なう。ここでの画像生成方式としては人物像生成部１３３で述べたような方式が利用できる。すなわち表示時間や表示像を変化させるための制御が行なわれ、静止画を用いる場合には複数の画像が用意され、動画を用いる場合には連続した動作の表示像の画像が指定される。また、コンピュータグラフィックスを利用する際には指定された動作の画像が生成される。

【０１１２】出力統合制御部１３７は、人物像生成部１３３の出力である表情を持つ人物像の画像情報、音声応答生成部１３５の出力である感情や強調を持つ音声応答の信号情報、応答文の文字列であるテキスト情報、内容可視化情報生成部１３６の出力である内容可視化情報を各部から受けとり、時間的な呈示順序を制御しながら、画面表示出力部１４と音声出力部１５に出力して利用者に応答内容を統合して呈示する。

【０１１３】ここで重要なことはそれぞれの出力を個別に呈示すれば良いのではなく、個々出力情報の時間的な関係を考慮しながら呈示する必要があることである。例えば、人物像が音声応答に合わせながら口を動作させる場合に音声応答出力と口動作の制御の同期やおじぎをしながら挨拶する場合の画像出力と音声出力の同期をとる必要がある。また、それぞれの出力の呈示順序を制御することが重要である。

【０１１４】図３０、３１、３２、３３に出力の呈示順序の時間制御の例を示す。図３０は最初の挨拶の場面の制御であり、まだ注文はないので、内容可視化情報は表示されないが、挨拶のテキスト情報をｔ０の時点で表示し、同時に人物像は「いらっしゃいませ」、続けて「ご注文をどうぞ」という音声応答に合わせて口を動作させながら、注文をうながす。このように発声している内容と人物像画面の同期をとり、あらかじめ分かり易いようにｔ０の時点でテキスト情報をすべて表示する。

【０１１５】図３１では既にハンバーガ１つとコーラ１つを注文済みの場面であり、応答確認内容の「ご注文はハンバーガを１つ、コーヒーを１つですね」のテキスト情報をｔ０の時点まで表示する。次いで人物像と音声応答を開始する時点のｔ０に内容可視化情報を新しく更新

しハンバーガ３つ、コーヒー２つ、コーラ１つを表示するようにする。また、人物像は音声の発声に合わせてｔ０からｔ３まで口を動かすように制御する。

【０１１６】この例で示した時間制御は音声応答の長さを基準に決められている。すなわち、図３０では「いらっしゃいませ」によりｔ０からｔ１まで、「ご注文をどうぞ」によりｔ１からｔ２までの継続時間が決まる。このそれぞれの継続時間は音声応答生成部１３５で決まるものであり、音声応答信号とその継続時間が出力統合制御部１３７に送られ時間制御に利用される。ここで述べた他にも呈示する内容可視化情報や人物像の画像の表示時間を基準に時間制御を行なうことも可能である。

【０１１７】図３２は、最初の注文を受けた後の全注文の確認の場面の制御であり、確認する品目はハンバーガ２つ、チーズバーガー１つ、コーヒー３つの３品目である。図３２では、応答内容の「ご注文はハンバーガが２つ、チーズバーガーが１つ、コーヒーが３つですね」のテキスト情報をｔ０の時刻で表示するとともに音声応答と、それに合わせた人物像の口の動作を開始する。音声の「ご注文は」までは、内容可視化情報の表示は行わないが、注文内容を発声し始めるｔ１の時点で内容可視化情報として、ハンバーガ２つ、チーズバーガー１つ、コーヒー３つを表示するようにする。また人物像は音声の発声に合わせてｔ０からｔ４まで口を動かすように制御する。

【０１１８】ここで、全注文の確認の応答文は応答文生成部１３１で生成されるが、確認する品目の数が多くなると生成される応答文は長くなり、音声応答の長さも長くなる。しかし、図３２の例において、利用者はｔ１の時点で表示される内容可視化情報により、音声応答を最後まで聞かずとも、システムの応答内容もしくはシステムの状態や理解内容を理解することができるため、内容可視化情報を表示した後のｔ１から音声応答が終わるｔ４までに出力される情報は利用者にとって冗長な応答である。

【０１１９】このため、本実施例では図３３に示すように、全注文の確認で、確認する品目が３つ以上ある場合は、出力の呈示順序を変え、最初のｔ０の時点で直前の応答文テキストを一端消去し、内容可視化情報として、注文内容であるハンバーガ２つ、チーズバーガー１つ、コーヒー３つを表示する。次に、この内容可視化情報の表示の処理が終ったｔ１の時点で「これでよろしいですか」という応答文テキストを表示するとともに、人物像と音声応答を開始する。この例で示した時間制御は、対話管理部１２で生成された出力意味表現のＡＣＴ情報と品目数をもとに出力統合制御１３７で行われ、応答文は、応答文生成部１３１で生成される。

【０１２０】またこれは、全注文の確認に限定されるものではなく、その他の確認の際に、応答確認内容の品目が多い場合や複雑でわかりづらい場合にも、最初に視覚

的応答出力を行った後、指示代名詞等を用いて短縮表現にした音声応答を行うことにより、対話を短時間に効率的に行うことも可能である。

【０１２１】尚、確認する品目数に代えて、他の音声応答の長さを示す指標、例えば音声応答中のワード数やモーラ数等、を用いて上述のような応答出力の変更を制御しても良い。

【０１２２】さらに、出力統合制御部１３７はそれぞれの画像表示情報の表示場所を制御している。例えば、画像出力装置１４の画面上で人物像を左上に、内容可視化情報を右に、テキスト情報を左下に制御し表示することが出来る。この表示位置は出力統合制御部１３７の制御の基に変更可能である。

【０１２３】以上のように本発明は、音声の入出力と画面表示の併用して対話を進めることを特徴としているが、ここで本発明における画面表示について実際の例を具体的に説明する。

【０１２４】まず、図３４は初期画面を示すもので、客が店頭にいない場合や近くに来ない場合には、「～へようこそ」など画面に文を表示するのみで、音声応答は出力しない。

【０１２５】ここで、ユーザ（客）がシステム（カウンターやドライブスルーの窓口等）に接近したような場合、例えば、圧力センサー付きのフロアーマットや監視カメラの画像等のセンサー情報によりユーザを検知すると、図３５に示すようにシステムは「いらっしゃいませ、御注文をどうぞ。」と漢字かな混り文で表示するとともに、ほほえみの表情の店員を画面上に表示して音声応答を出力する（図９の状態Ｓ０）。

【０１２６】この時、ユーザの検知は、人の動きや位置を考慮し、特に、立ち止まりを検出した時点で上記の処理を実行し安心してユーザとの音声対話をスタートさせることが重要である。特に、店員の笑顔は、客をリラックスさせる効果があり、明るい声を合成することも望ましい。これらの技術はすでに開発されており、また、録音された合成音や自然音声を用いることも可能である。

【０１２７】この状態から、ユーザが画面を見ながら、仮に早口で注文を「え～、ハンバーガを２つとあの～コーヒーを２つお願いします。あ～」と行なったとする。すると、システムでは、図９の状態Ｕ０のユーザの発音を処理するが、いま聞きとれない部分があり、図１に示す音声理解部１１から何の結果も得られないとすると、対話管理部１２ではリジェクトに対応する。

【０１２８】この場合、図３６に示すようにシステムは「はっきり聞きとれませんでした。もう一度お願いします。」と漢字かな混じり文を表示するとともに、申し訳なさそうな表情の店員を画面上に表示して音声応答する。この状態では、システム側は、ユーザの注文を全く聞きとれず、その時点の注文テーブルは何もない（空）状態なので、注文に関する画面表示は何もなされない。

また、店員の表情生成は、応答文の関与として出力される。この場合、図９のユーザの状態Ｕ０から、リジェクト対話修正状態Ｓ１０に移し、ここで応答と表情が決定されることになる。

【０１２９】次に、このシステムからの応答を受けとったユーザが、前回よりもはっきりとした話し方で「ハンバーガ２つとコーヒー２つ下さい。」と注文を行なったとすると、前述した音声理解処理、対話処理が実行され、入力意味表現と注文テーブルを生成した後、出力意味表現が決定される。そして、出力意味の応答ＡＣＴが“全確認”となると、次の応答として図３７に示す画面表示と音声応答が併用して行なわれる。

【０１３０】この場合、システムは「御注文は、ハンバーガ２つ、コーヒーを２つですね。」と漢字かな混り文で表示するとともに、店員の顔を画面上に表示して音声応答を出力するようになる。この時の店員の表情と音声応答の感情については、前述したように文と状態を考慮して決定され、ここでは普通の表情と音声応答が出力される。また、応答文とともに、注文テーブルの内容が画面表示され、ユーザは自分の要求した品物かどうか、個数かを短時間で確認するようになる。

【０１３１】この場合、品物の表示は、個数を数字で現さず品物を注文個数だけ並べた状態を画面表示してもかまわない。ここでは数字の情報が重要なのでハンバーガ等の品物と同じ高さの領域を設けて数字を表示している。すなわち、個数（数字）についての情報は重要であり、ユーザにそのことを自然に伝えられるようにしている。また、数字の表示サイズについても、大きさで大体の情報が伝えられるので、大きさを変えて表示することも有効であり、また、文字情報を併用したり、カラー情報などを併用して出力することにより、音声応答やテキスト応答よりもリアルなイメージを自然に素早くユーザに伝えることが可能となり、より高速な確認対話を実現している。一方、店員に関する人物像については、リアルな表情よりも、伝えたいポイントが伝わる情報量のすくない絵が有効である。また、上記の画像表示は、三次元グラフィックスを用いても当然に行なうことができる。

【０１３２】さて、システム側からの注文品を確認されてユーザが「え～と、まあ、それでいいや」と少し迷いながら発音したとする。すると、システムでは、音声理解部１１からの何の結果も得られないことで、対話管理部１２がリジェクトに対応する。この場合、図３８に示すようにシステムは「すみません。もう一度入力して下さい。」と漢字かな混り文を表示するとともに、店員を画面上に表示して音声応答する。この場合のメッセージは、上述した図３６の場合よりも手短なものであり、音声対話を手短に伝えるように状態と対話の履歴情報を用いて応答文が決定される。また、店員の顔の表情についても応答文に対応して、申し訳なさそうなものが出力される。

【０１３３】この画面表示のポイントは、現状でシステム側が理解している注文の内容を右側の領域に表示している点である。この注文品の表示は、注文テーブルをそのまま表示するので、音声が持つ一過性の欠点を補うことができる。すなわち、追加や置換や削除についての確認は音声応答、応答文で一部分の局所的な注文について行なうが、対話の進行に伴う蓄積された注文、確認の結果である注文品の表示による効果は大きい。

【０１３４】そして、このような表示は、前述した対話管理部１２での処理で容易に実現できる。また、部分確認に視覚表示を用いることも可能であり、注文品の全表示を続けて別の表示領域で行なうこともできる。さらに、全注文品の表示を一時的に隠し、部分確認にユーザの意識を集中させるために、部分確認の画面表示を行なうこともできる。すなわち、両者の長所を組み合わせた表示方法の併用を状況により使い分けて行なうことが効果的である。

【０１３５】この後、ユーザがはっきりした声で「それで、いいです。」と発声すると、システムはこれを肯定と理解して図９のＳ９に遷移し、図３９に示すようにシステムは「ありがとうございました」と漢字かな混り文を表示するとともに、頭を下げた店員を画面上に表示して音声応答し、対話を終了する。

【０１３６】この時の応答文生成、笑顔の生成、おじぎをするジェスチャーの生成も、上述したのと同様の処理により行なう。また、合計金額の表示なども、種々の応答（音声、画面表示）で行なえる。

【０１３７】なお、上述した図３８の確認の場合、図４０に示すように「はい」、「いいえ」と答を誘導するように画面表示を行なうのも効果的である。この場合、聞き返しや訂正の回数の情報が使え、システムは「すみません。御注文はハンバーガを２つ、コーヒーを２つですか。はいか、いいえでお答え下さい。」のような状況に応じた対話が行える利点がある。

【０１３８】図４１は、このような実施例での処理手順を簡単にまとめたものである。

【０１３９】この場合、フロアマット２２０がユーザを検知すると、ユーザからの音声入力についてキーワード検出部２１によりキーワードを検出し、単語候補系列２２２を求め、次いで、構文意味解析部２２でキーワードに基づく音声理解を行ない、入力意味表現２２４を求める。そして、さらに対話制御部１２で対話と応用分野の知識による対話処理が行なわれ、出力意味表現２２６を求め、これを応答生成部１３に与えて、ここでの規則合成による音声応答出力と画面表示出力とから成るマルチモーダルは応答を行う。

【０１４０】以上の説明は、ファースト・フードの注文の例で行なったが、情報サービスやマルチメディア、ワークステーションおよび通信ネットワークを用いた座席

予約システムなどへの運用も可能である。

【０１４１】次に、本発明の他の実施例を図４２により説明する。

【０１４２】図４２は、本発明の音声対話システムに人の動き状態を検出する機能を組み込んだものを示している。この場合、人状態検出は、システムが対話を自動的に始め、そして終了させるのに必要な機能で、対話の開始、終了におけるユーザの状態や反応を理解することで、対話を自然に進めることを可能とするものである。人状態検出の方法としては、光、超音波、赤外線圧力などを処理して行うことが考えられるが、ここでは、大人一人を検出できるフロアマットを利用した例について述べる。

【０１４３】図４２では、図１で述べたと同様な音声入力部２３１、音声理解部２３２、対話管理部２３４、応答生成部２３５、ディスプレイ２３６、スピーカ２３７の他に人状態検出部２３３を設ける構成になっている。

【０１４４】この場合、人状態検出部２３３は、図４３に示すようにマットに人が乗っている場合には人状態検出意味表現１を、マット上から人が降りた状態には人状態検出意味表現２を出力するようになっていて、これらの出力を対話管理部２３４に通知するようにしている。

【０１４５】対話管理部２３４は、人状態検出部２３３からの人状態検出意味表現の他に、上述した実施例と同様に音声理解部２３２からも入力意味表現を取り込み、対話の知識や対話の履歴情報を用いて確認応答のための出力意味表現を生成する。

【０１４６】この場合、対話管理部２３４では音声理解部２３２からの入力意味表現と人状態検出部２３３からの人状態検出意味表現を受け取る際に、対話の状態によりそれぞれの意味表現を順に処理したり、優先的に処理することができ、ユーザの状態や各種の反応を理解し対話を進めることができるようになっている。

【０１４７】しかして、ユーザがマットに乗ると人状態検出部２３３より人状態検出意味表現１が出力され対話管理部２３４に送られる。すると、対話管理部２３４より挨拶の出力意味表現１が応答生成出力部２３５に送られ、応答出力として「いらっしゃいませ、ご注文をどうぞ」がディスプレイ２３６およびスピーカ２３７より出力される。

【０１４８】次に、ユーザが「ハンバーガとコーヒー２つづつ」と入力すると、音声理解部２３２より入力意味表現１が出力され対話管理部２３４に送られる。これにより対話管理部２３４では、入力意味表現と注文テーブルの内容を参照し、出力意味表現２を出力し、応答生成出力部２３５を通して「ハンバーガ２こにコーヒー２こですね」の応答が出力されるようになる。

【０１４９】この場合、通常は、図４４に示すように「ハンバーガ２こコーヒー２こですね」「はい」「ありがとうございました。」というように対話が進んでいくが、ユーザが途中でマット上から離れてしまったような場合は図４５のようになる。

【０１５０】すなわち、出力意味表現２の「ハンバーガ２こにコーヒー２こですね」の応答が出力された後で、人状態検出部２３３より人状態検出意味表現２が出力され、対話管理部２３４に入力されるようになる。この場合は、対話管理部２３４は発話内容の確認を行なわれずにユーザが立ち去ったことから、注文内容の登録は行なわずに、出力意味表現４の「ご利用ありがとうございました」という自然な応答を出力するようになる。

【０１５１】このようにして、人状態検出部２３３を対話管理部２３４と組み合わせることにより、ユーザの状態や反応を理解することが可能となり、自然に対話を進めることができる。

【０１５２】なお、本実施例では人の状態検出にマットを用いたが、これに限られるものではなく、監視カメラなどの他の方法を用いてもよい。

【０１５３】次に、このような処理を図４６のフローチャートにより説明する。

【０１５４】この場合、システムは対話管理２３４において状態（ｓｔａｔｅ）＃０，＃１，＃２，＃３を持ち、初期状態は状態＃０である（ステップＳ２８１）。状態＃０においては人状態検出意味表現の人状態ＡＣＴが「人存在」であるかを確認し（ステップＳ２８２）、人がいる場合には状態を＃１にし、挨拶の出力意味表現により応答を生成し出力するようになる（ステップＳ２８３）。

【０１５５】次に、状態＃１において、音声理解部２３２から入力意味表現の発話ＡＣＴが注文の場合は（ステップＳ２８４，Ｓ２８５）、対話知識に基づいて注文内容の確認の出力意味表現を送出し応答を出力するようになる。また、発話ＡＣＴがはいの場合は（ステップＳ２８７）、状態を＃２にするとともに、発話アクトはいに対応する出力意味表現により応答を出力するようになる（ステップＳ２８８）。また、発話ＡＣＴがいいえの場合は（ステップＳ２８９）、注文内容の再確認の出力意味表現を送出し応答を出力するようになる。さらに、人状態検出意味表現の人状態ＡＣＴが「人不在」であることを確認した場合は（ステップＳ２９１）、状態を＃３にする。

【０１５６】そして、状態＃２においては、お礼１として「ありがとうございました」を出力し（ステップＳ２９３，Ｓ２９４）、状態＃３においては、お礼２として「ご利用ありがとうございました」を出力するようになる（ステップＳ２９５，Ｓ２９６）。

【０１５７】次に、本発明の他の実施例を図４７により説明する。

【０１５８】この実施例は、図１で述べた音声理解部１１、応答生成出力部１３での音声入出力、人状態検出を行う部分について詳述するものである。

【０１５９】この場合、音声対話システム全体は図４７に示すように演算部２９１、メモリ部２９２、保存部２９３、保存部インターフェース２９３１、通信部２９４、通信部インターフェース２９４１、Ａ／Ｄ部２９５、マット部２９６、演算処理部２９７、Ｄ／Ａ部２９８、表示部２９９から構成されている。

【０１６０】ここで、Ａ／Ｄ部２９５は、マイク２９５１、フィルタ増幅部２９５２、Ａ／Ｄ変換部２９５３、Ａ／Ｄ変換部インターフェース２９５４からなっている。フィルタ増幅部２９５２は、マイク２９５１からの入力の増幅およびＡ／Ｄ変換のための高域遮断フィルタ機能を有している。ここでのフィルタの遮断周波数は、Ａ／Ｄ変換のサンプリング周波数で決まるが、例えば１２ｋＨｚでサンプリングする場合には、５．４ｋＨｚで高域周波数成分を遮断するようになる。また、Ａ／Ｄ変換部２９５３は増幅された入力音声を、例えば１６ｋＨｚ又は１２ｋＨｚでデジタル化し、Ａ／Ｄ変換部インターフェース２９５４内に一時保存するとともに、演算部２９１の制御によりメモリ部２９２に転送するようにしている。

【０１６１】マット部２９６はマット２９６１、マット制御部２９６２、マット制御部インターフェース２９６３からなっていて、マット２９６１上での人の存在／不在をマット制御部２９６２で検出し、この結果をマット制御部インターフェース２９６３を通じて転送するようにしている。

【０１６２】演算処理部２９７は、高速演算処理部２９７１、高速演算処理部インターフェース２９７２からなっている。高速演算処理部２９７１は音声理解処理、応答生成処理さらには画像処理による人状態検出処理などの大量な演算に必要な処理に使用する。この場合、このような処理は、同時に処理する必要があり、複数の高速演算処理部２９７１を同時に使用できるようにしている。また、それぞれの演算処理は、演算部２９１の制御の下に入力データをメモリ部２９２から高速演算処理部２９７１に転送し、処理終了後に結果をメモリ部２９２に転送するようにしている。

【０１６３】Ｄ／Ａ部２９８はＤ／Ａ変換部インターフェース２９８１、Ｄ／Ａ変換部２９８２、フィルタ増幅部２９８３、スピーカ２９８４からなり、演算部２９１の制御の下でメモリ部２９２に記憶されたデジタルデータをＤ／Ａ変換部インターフェース２９８１を通じてＤ／Ａ変換部２９８２に転送し、ここで一定周期、例えば１２ｋＨｚでアナログデータに変換し、フィルタ増幅部２９８３を通してスピーカ２９８４に出力するようにしている。この場合、Ｄ／Ａ変換部２９８２はデータの一時保存部を有し、メモリ部２９２からのデータ転送を高速に行うことで、演算部２９１が他の処理も行うことができるようにしている。

【０１６４】表示部２９９は表示制御部インターフェース２９９１、表示制御部２９９２、ディスプレイ２９９３からなり、演算部２９１の制御の下で画像、文字、図形、動画情報、色や輝度、濃度情報の変化などの情報を表示制御部２９９２よりディスプレイ２９９３に表示するようにしている。

【０１６５】通信部２９４は、外部の計算機、情報処理器、サービス処理機器などと制御情報データの通信を行うもので、各データは演算部２９１の制御により通信部インターフェース２９４１を通じてやり取りされる。

【０１６６】保存部２９３は、演算部２９１の制御の下に音声理解、対話管理、応答生成に必要なデータ、制御情報、プログラム、中間情報などを保存している。

【０１６７】演算部２９１はメモリ部２９２に記憶された各部の情報、実行プログラム、そのためのプログラムを使用してＡ／Ｄ部２９５、マット部２９６、演算処理部２９７、Ｄ／Ａ部２９８、通信部２９４、保存部２９３の制御を行うようにしている。

【０１６８】ここで、演算部２９１が実行するプログラムは、図１で述べた音声理解部１１、対話管理部１２、応答生成出力部１３での処理を行い、マルチタスクの形式で実行される。そのためのタスクの切り替えは、一定時間ごとに順次行われるが、各部の処理や入出力が完了した場合など、処理を優先させる必要がある場合には、割り込みにより、その処理を優先させる。

【０１６９】上述ではＡ／Ｄ部２９５、Ｄ／Ａ部２９８については、それぞれ個別に動作できるようにしている。これにより、音声入力、合成音出力を同時に、しかも別々に取り扱うことができるので、合成音出力中でも音声入力が可能となり、合成音キャンセルにより入力音声の検出および認識が可能になる。

【０１７０】しかし、これらＡ／Ｄ部２９５、Ｄ／Ａ部２９８の構成として、図４８（ａ）に示すように共通のＡ／Ｄ、Ｄ／Ａ変換部インターフェース３０１を用いるようにしたり、図４８（ｂ）に示すように共通のＡ／Ｄ、Ｄ／Ａ変換部インターフェース３０２、Ａ／Ｄ、Ｄ／Ａ変換フィルタ部３０３および増幅部３０４を用いるようにしてもよい。

【０１７１】ところが、このような構成では、データのやり取りを同時に双方向でできず、入力か出力のどちらかに限られてしまうため、合成音出力中の音声入力の受付けと同時に合成音を出力することができない。

【０１７２】この場合、ユーザは音声入力の受付け状態を知る手段がないときに、受付けられていない発話についての応答を待ったり、発話の前半が入力されなかったりする不都合が生じる。そこで、音声の入出力許可状態を画像表示することによりユーザに対して計算機側が音声の入出力許可状況を伝えることができる。特に、画像表示と文字表示を組み合わせることで、例えば、図４９に示すように「くちびる」と「ＳＰＥＡＫ」により発声できる状態、図５０に示すように封止された「くちび

る」と「ＬＩＳＴＥＮ」により発声できない状態をそれぞれ表示することができる。

【０１７３】このように各入出力機器の状態、状況を別の出力機器により伝えることができ、より自然で分かりやすい対話が可能になる。さらに、各入出力機器の状況だけでなく、ユーザに対し重要なことで注意して聞いてもらう必要がある場合や対話管理において音声入力を行ってほしくない場合などにも画像表示と文字表示の組み合わせや、さらに色や輝度、濃度の変化などにより注意を促すことができる。

【０１７４】本発明は、上記実施例にのみ限定されず、要旨を変更しない範囲で、適宜変形して実施できる。

【０１７５】

【発明の効果】以上説明したように、本発明では、システム側からユーザへ応答を出力する際に、音声応答の他に、人物（の顔）を表示システムに対する親近感を持たせると同時に音声応答と同期して口を動かし、ユーザの注目点を定め、使い勝手を向上させている。また、同一画面上に、音声応答の品質の低さをカバーするため音声応答文もテキスト・データの形で表示し、音声の発話速度よりも速くユーザは応答文を受けとることを可能とする。さらに、同一画面上に応答内容を視覚化（Ｖｉｓｕａｌｉｚａｔｉｏｎ）したものを表示し、種々の応用に適した形態で伝達すべき重要なメッセージの意味や内容を、人物表示音声応答や音声応答文と同期させて出力することにより、ユーザが一見してわかるようなシステム側からユーザへの高速なメッセージの伝達が実現できる。

【０１７６】以上のように種々の形態の応答をシステム側から同一画面上に並行にユーザに出力するため、ユーザは状況に応じて適当な個々の応答を選択したり、２種類、あるいは、３種類の形態の応答データを同時に受けとることが可能となり各メディアの有する特徴を活かすという効果が得られ、ユーザにとっての自由度が増し、使い勝手のよいマルチモーダルなヒューマンインターフェースが実現できる。

【０１７７】この結果、従来問題であった音声対話システムの音声確認部の誤確認や曖昧性に基づく不完全さを、対話によりスピーディに効率的にカバーし、対話の進行により、ユーザの意図していることの理解が容易になる。

【０１７８】また、入力側にマットやカメラ等による人状態検出手段を設け、ユーザ検出とともに、単に合成音を出すだけではなく表示画面上の人物の顔を明るくし、笑顔にするようにもできるので、対話のタイミングが良くなるばかりでなく、ユーザが驚かずに安心して使えるユーザフレンドリーな音声インターフェースが実現できる。さらに、マルチモーダル音声対話システムに適用することにより、使い勝手の良い自然なシステムが得られ、コンピュータの操作性が著しく向上するという効果が得られる。

【０１７９】また、音声応答のキャンセル機能を加えることにより、音声応答中にでも画面表示結果をユーザがみて、常時、割り込む（Ｉｎｔｅｒｒｕｐｔ）ことが可能になり、スピーディーな音声による対話が可能であり、音声認識性能が低い場合でも対話のやり取りでカバーし、意図の伝達やデータ入力の能率を大幅に改善できる。

【０１８０】以上を総合すると本発明では、音声認識と音声合成と対話管理機能を具備する音声対話システムにおいて、システム側からユーザへの応答に際して時系列情報である音声応答と並行して、応答内容の可視化を行い、特に、表情やジェスチャーの表示、品物やサイズ、種別等の対償物（オブジェクト）の表示、応答文の文字出力を並行して行うことができることから、ユーザは同時に様々な観点から応答を受け取ることがてきるようになり、自由度が増し、必要に応じて正確な情報を選択でき、親しみ易さ、効率、快適さの改善、目や耳の疲労度の軽減等に効果的である。

【図面の簡単な説明】

【図１】本発明の一実施例の概略構成を示す図。

【図２】音声理解部の詳細構成を示す図。

【図３】キーワード候補系列を説明するための図。

【図４】入力意味表現の一例を示す図。

【図５】キーワードの内容を示す図。

【図６】出力意味表現の一例を示す図。

【図７】注文テーブルの一例を示す図。

【図８】旧注文テーブルの一例を示す図。

【図９】対話管理部の内部における状態遷移の一例を示す図。

【図１０】ユーザ状態の処理を説明するためのフローチャート。

【図１１】システム状態の処理を説明するためのフローチャート。

【図１２】対話処理の具体的な例を示す図。

【図１３】図１２に示す対話処理におけるユーザ状態の処理を説明するための図。

【図１４】対話処理におけるシステム状態の処理を説明するための図。

【図１５】図１４における対話処理の具体的な例を示す図。

【図１６】システムからの出力応答文の例を示す図。

【図１７】ユーザの発声速度の求め方を説明する図。

【図１８】対話管理部の出力の一例を示す図。

【図１９】応答生成出力部の詳細構成を示す図。

【図２０】人物像情報の一例を示す図。

【図２１】応答文型の例を示す図。

【図２２】応答文生成部における応答文の生成のフローチャートと具体例を示す図。

【図２３】人物像表情決定部の一例を示す図。

【図２４】感情・強調決定部の一例を示す図。
【図２５】音声応答生成部の詳細構成を示す図。
【図２６】基本周波数パターンモデルの一例を示す図。
【図２７】喜びを併う応答の場合の基本周波数パターンの変化を示す図。
【図２８】悲しみを併う応答の場合の基本周波数パターンの変化を示す図。
【図２９】音声波形生成部の具体的構成の一例を示す図。
【図３０】出力呈示順序の時間制御の例を示すタイミングチャート。
【図３１】出力呈示順序の時間制御の他の例を示すタイミングチャート。
【図３２】出力呈示順序の時間制御の他の例を示すタイミングチャート。
【図３３】出力呈示順序の時間制御の他の例を示すタイミングチャート。
【図３４】表示画面での表示例を示す図。
【図３５】表示画面での表示例を示す図。
【図３６】表示画面での表示例を示す図。
【図３７】表示画面での表示例を示す図。
【図３８】表示画面での表示例を示す図。
【図３９】表示画面での表示例を示す図。
【図４０】表示画面での表示例を示す図。
【図４１】対話処理の手順を簡単にまとめて示す図。
【図４２】本発明の他の実施例の概略構成を示す図。
【図４３】人状態検出部を説明するための図。
【図４４】図４２に示す他の実施例の動作を説明するための図。
【図４５】図４２に示す他の実施例の動作を説明するための図。
【図４６】図４２に示す他の実施例の動作を説明するた＊めのフローチャート。
【図４７】本発明の他の実施例の概略構成を示す図。
【図４８】図４７に示す他の実施例の一部を変形した例を示す図。
【図４９】表示画面での表示例を示す図。
【図５０】表示画面での表示例を示す図。

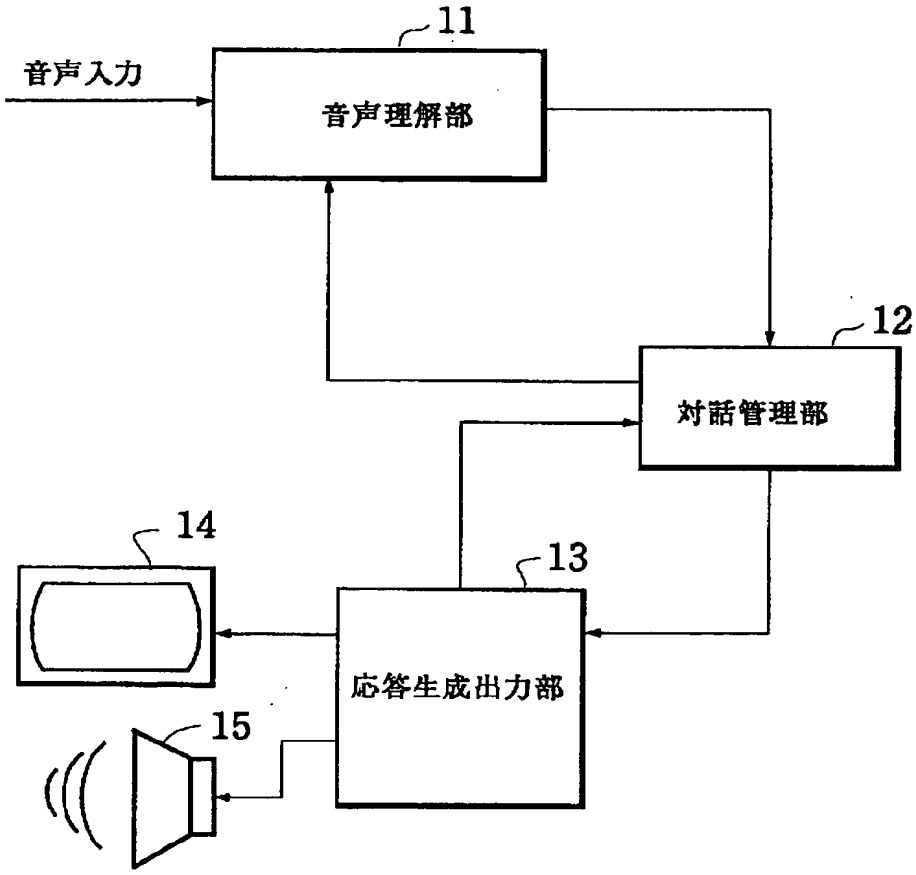
【符号の説明】

１１，２３２　音声理解部
１２，２３４　対話管理部
１３，２３５　応答生成部
１４，２３６　ディスプレイ
１５，２３７　スピーカ
２１　キーワード検出部
２１ａ　音声分析部
２１ｂ　キーワードスポッティング処理部
２２　構文意味解析部
２２ａ　文始端判定部
２２ｂ　文候補解析部
２２ｃ　文終端判定部
２２ｄ　文候補テーブル
２３１　音声入力部
２３３　人状態検出部
２９１　演算部
２９２　メモリ部
２９３　保存部
２９４　通信部
２９５　Ａ／Ｄ部
２９６　マット部
２９７　演算処理部
２９８　Ｄ／Ａ部
２９９　表示部

【図６】

ACT ＝ 確認		
品名	サイズ	個数
ハンバーガー	0	1
コーヒー	0	2
ポテト	大	4

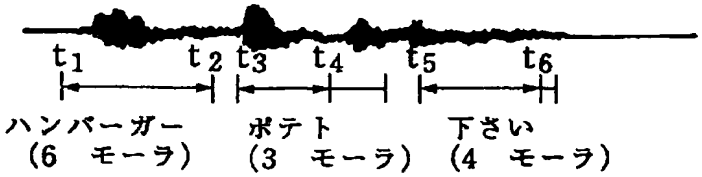
【図１】



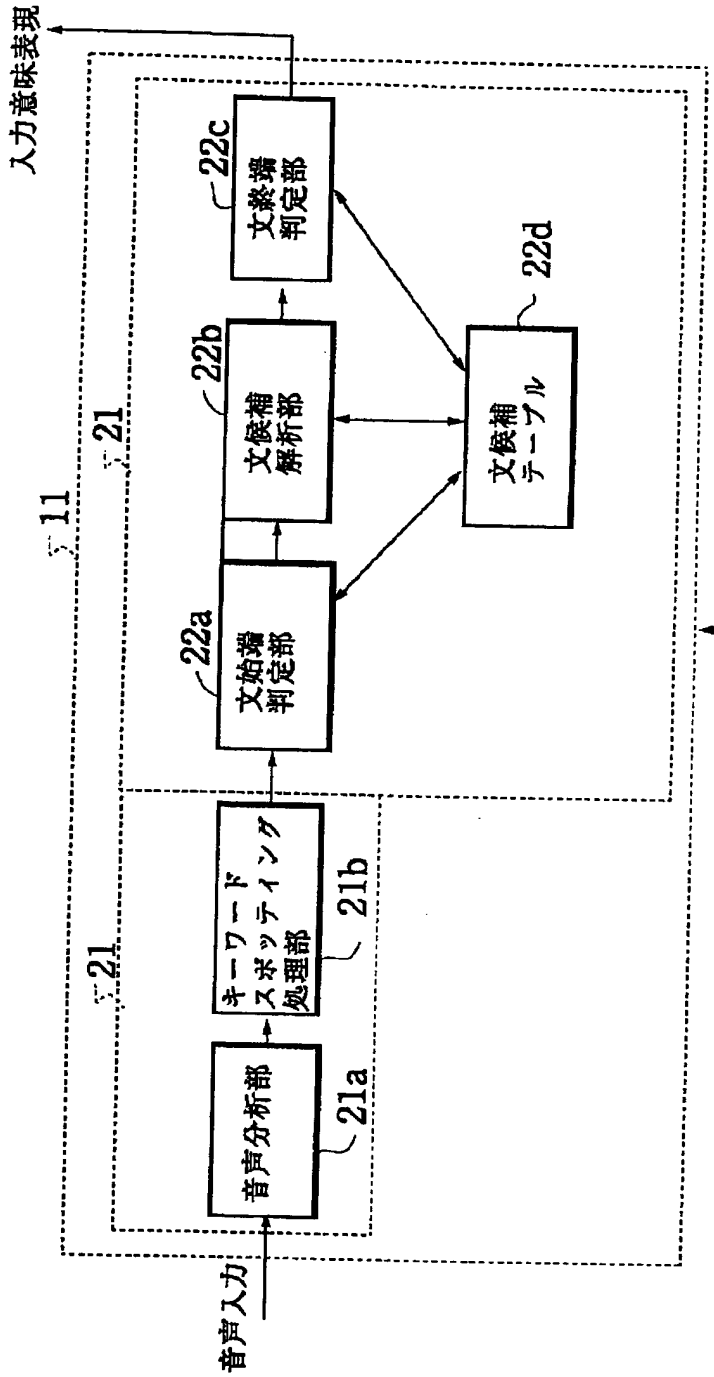
【図７】

品名	サイズ	個数
ハンバーガー	0	2
ポテト	大	2

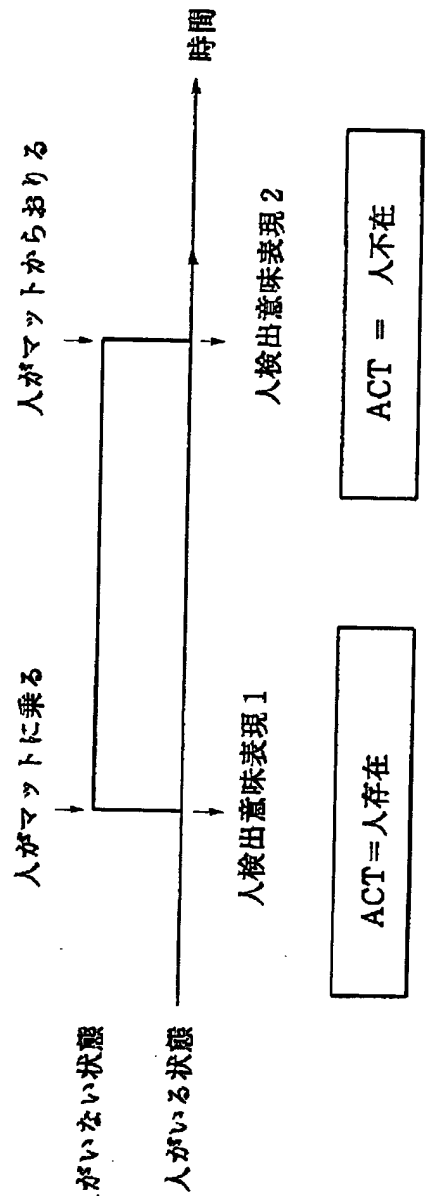
【図１７】



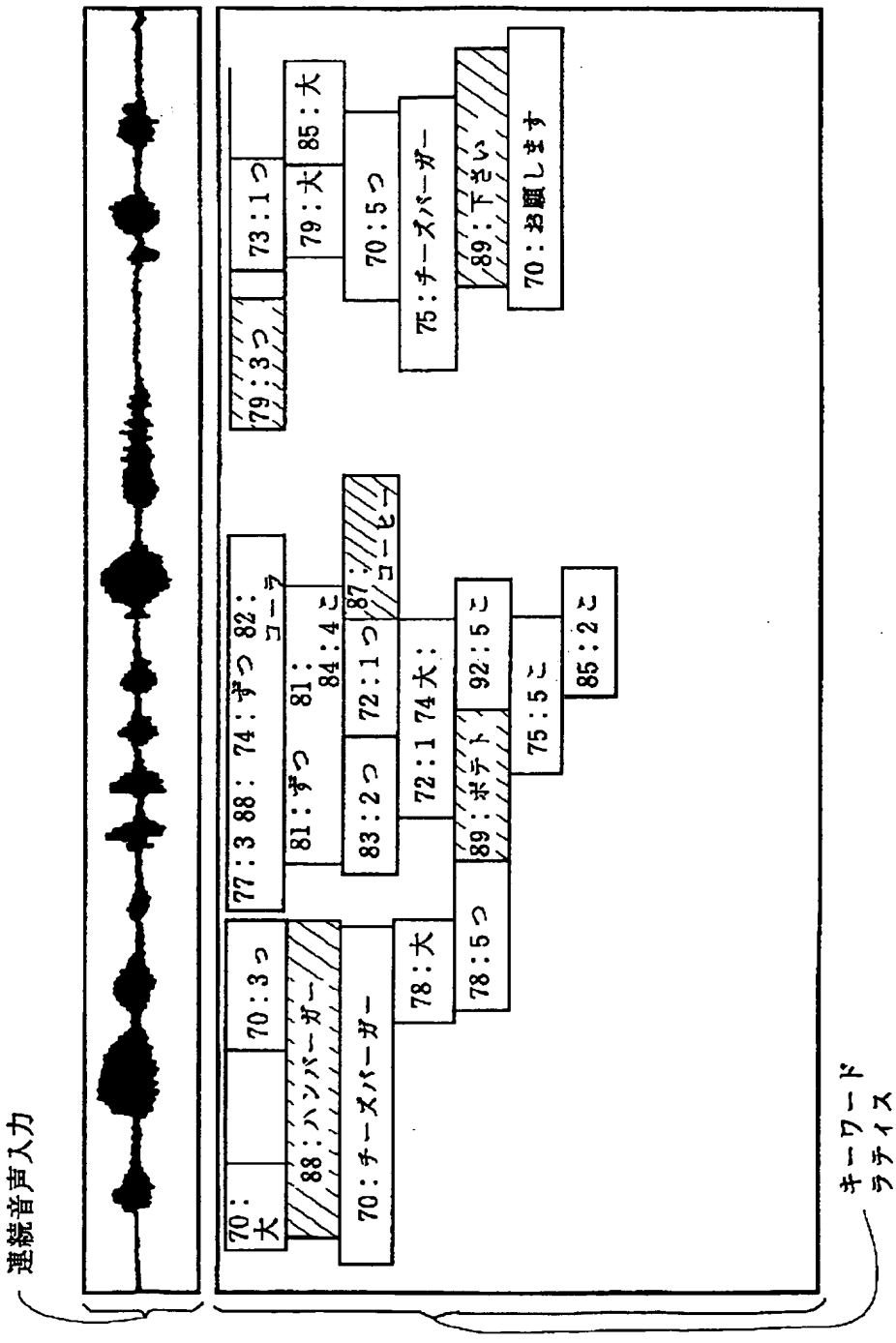
【図2】


入力意味表現
11
21
21
22c
文終端判定部
22b
文候補解析部
22d
文候補テーブル
22a
文始端判定部
対話管理部12から
キーワードスポッティング処理部
21b
音声分析部
21a
音声入力

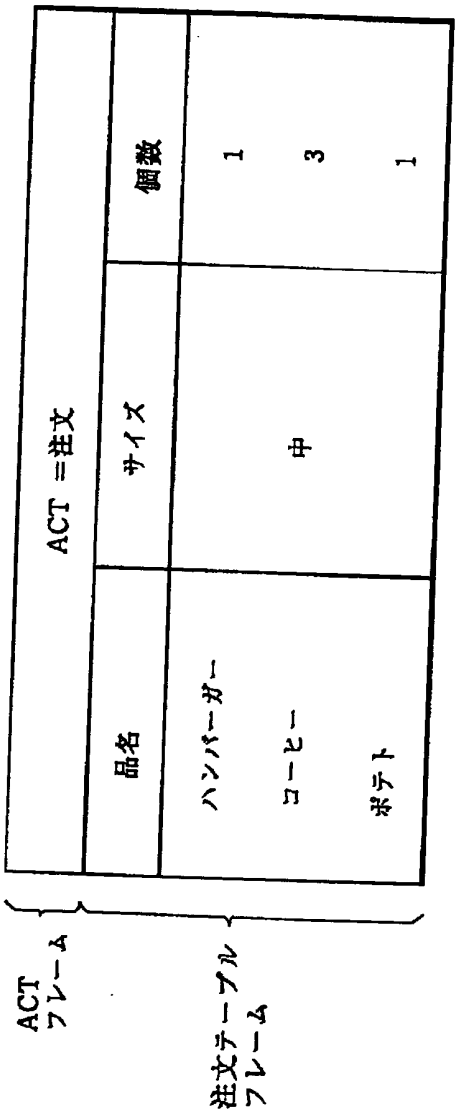
【図43】


時間
人がマットからおりる
人がマットに乗る
人検出意味表現2
人検出意味表現1
人がいない状態
人がいる状態
ACT＝人存在
ACT ＝ 人不在

【図３】


連続音声入力
キーワード
ラティス
70：大
88：ハンバーガー
70：チーズバーガー
70：3っ
78：大
78：5つ
77：3
88：
74：ずつ
82：コーラ
81：ずつ
81：84：4こ
83：2つ
72：1つ
87：コーヒー
72：1
74大：
89：ポテト
92：5こ
75：5こ
85：2こ
79：3つ
73：1つ
79：大
85：大
70：5つ
75：チーズバーガー
89：下さい
70：お願します

【図４】



【図5】

1. ハンバーガー
2. チーズバーガー
3. フィッシュバーガー
4. ポテト
5. フライドポテト
6. コーヒー
7. アイスコーヒー
8. コーラー
9. オレンジジュース
10. 一個
11. 二個
12. 三個
13. 四個
14. 五個
15. 一つ(ひとつ)
16. 二つ(ふたつ)
17. 三つ(みっつ)
18. 四つ(よっつ)
19. 五つ(いつつ)
20. はい
21. ええ
22. そうです
23. いーです
24. いいえ
25. 違います
26. 違う
27. 要らない
28. 要りません
29. 取消し
30. あと
31. それから
32. それと
33. 追加
34. ではなくて
35. じゃなくて
36. やめて
37. 下さい
38. お願いします
39. ちょーだい
40. づつ
41. 全部
42. それぞれ
43. みんな
44. 大(だい)
45. 中(ちゅー)
46. 小(しょー)
47. 大きい(おーきい)
48. 普通(ふつー)
49. 小さい(ちーさい)

【図15】

(a)

ACT＝追加		
品名	サイズ	個数
コーラ	大	1

(b)

応答リスト
全体確認 部分確認 個別確認 追加確認 削除確認 置換確認

(c)

ACT＝追加確認		
品名	サイズ	個数
コーラ	大	1

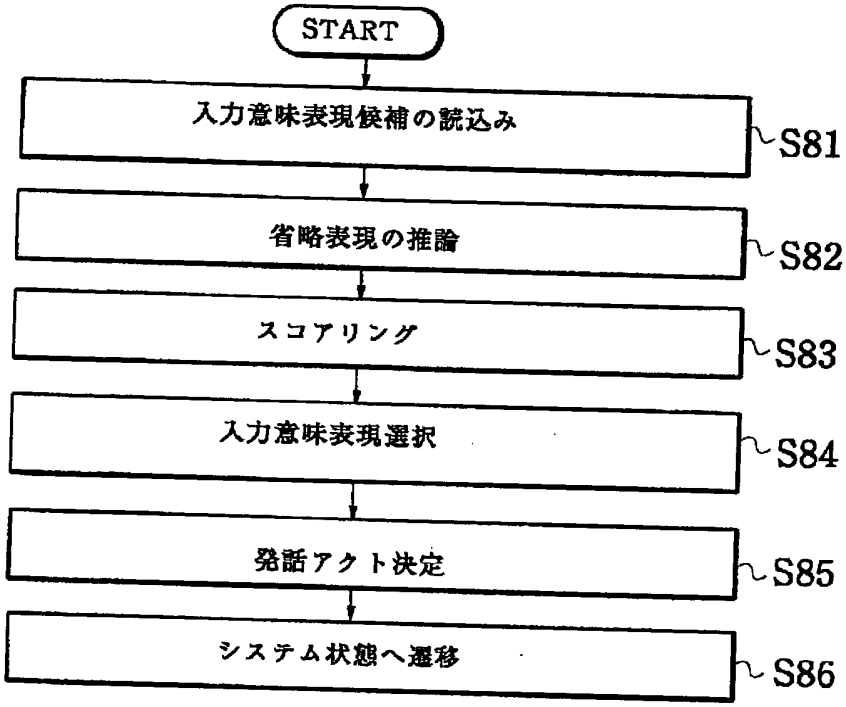
【図20】

システム状態	SP 1
ユーザ状態	UP 3
繰返し回数 N	0
強調項目	なし
確信度 D	1.0

【図8】

品名	サイズ	個数
ハンバーガー	0	1

【図10】



【図21】

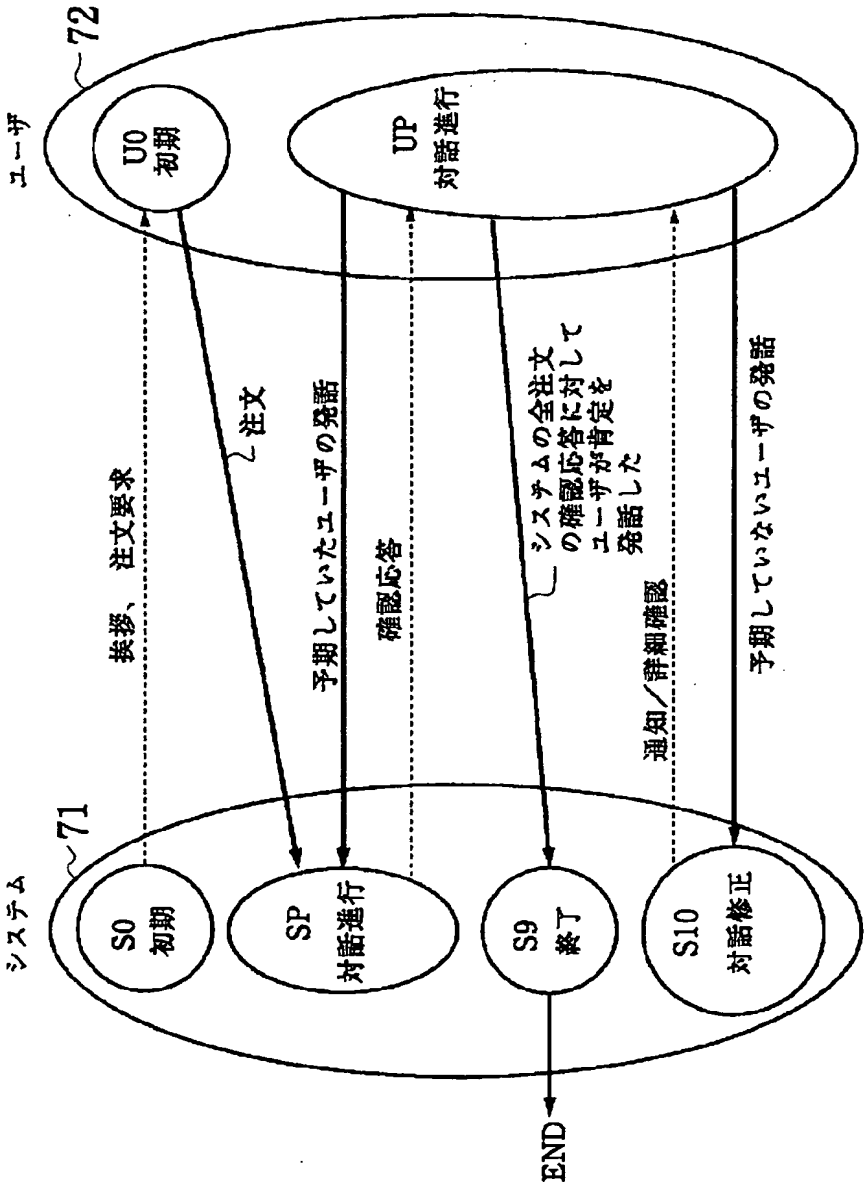
ACT ＝ 部分確認

「確認します。〈品名〉〈サイズ〉〈個数〉 ですね。」

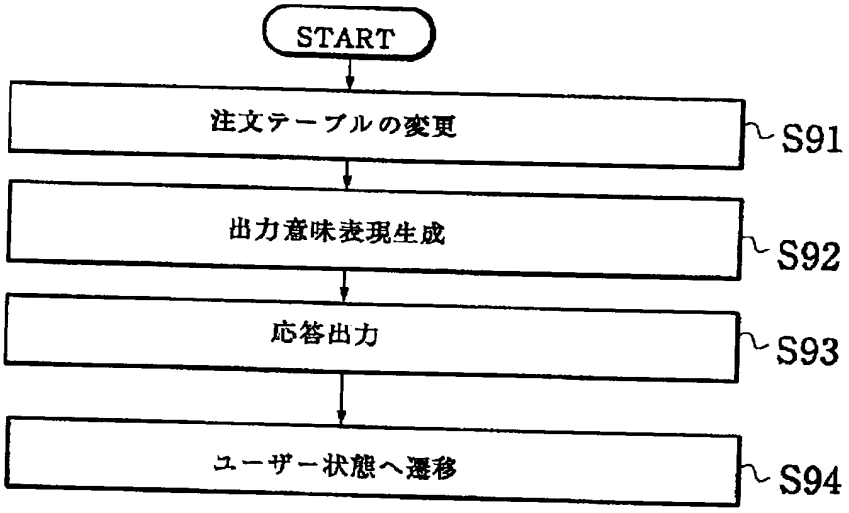
ACT ＝ 追加確認

「確認します。〈品名〉〈サイズ〉〈個数〉 追加ですね。」

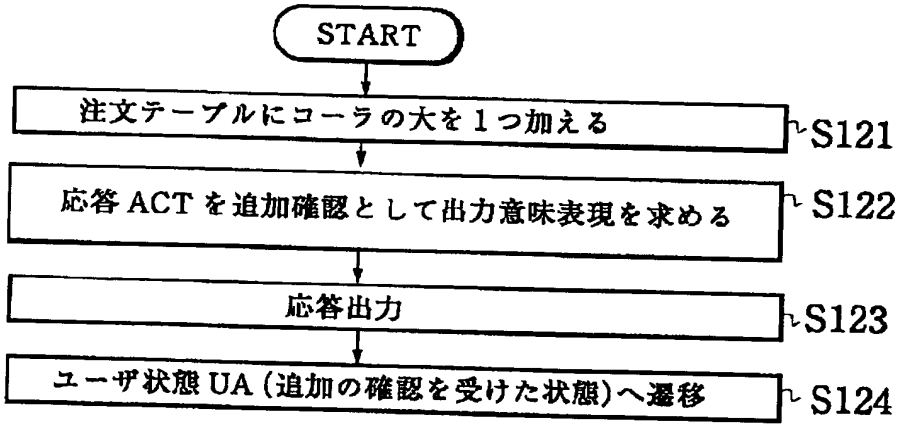
【図9】


72
ユーザ
U0
初期
UP
対話進行
挨拶、注文要求
注文
予期していたユーザの発話
確認応答
システムの全注文
の確認応答に対して
ユーザが肯定を
発話した
通知／詳細確認
予期していないユーザの発話
71
システム
S0
初期
SP
対話進行
S9
終了
S10
対話修正
END

【図１１】


START
注文テーブルの変更
S91
出力意味表現生成
S92
応答出力
S93
ユーザー状態へ遷移
S94

【図１４】


START
注文テーブルにコーラの大を１つ加える
S121
応答 ACT を追加確認として出力意味表現を求める
S122
応答出力
S123
ユーザ状態 UA（追加の確認を受けた状態）へ遷移
S124

【図１２】

(a)出力意味表現

ACT＝注文		
品名	サイズ	個数
ハンバーガー	0	1
コーヒー	0	2
コーラ	大	4

(b)注文テーブル

品名	サイズ	個数
ハンバーガー	0	1
コーヒー	0	2
コーラ	大	4

(c)

システム　「御注文は、ハンバーガーを１つ
　　コーヒーを２つ
　　コーラの大を４つ
ですね。」

ユーザ　「コーラを１つ追加して下さい。」

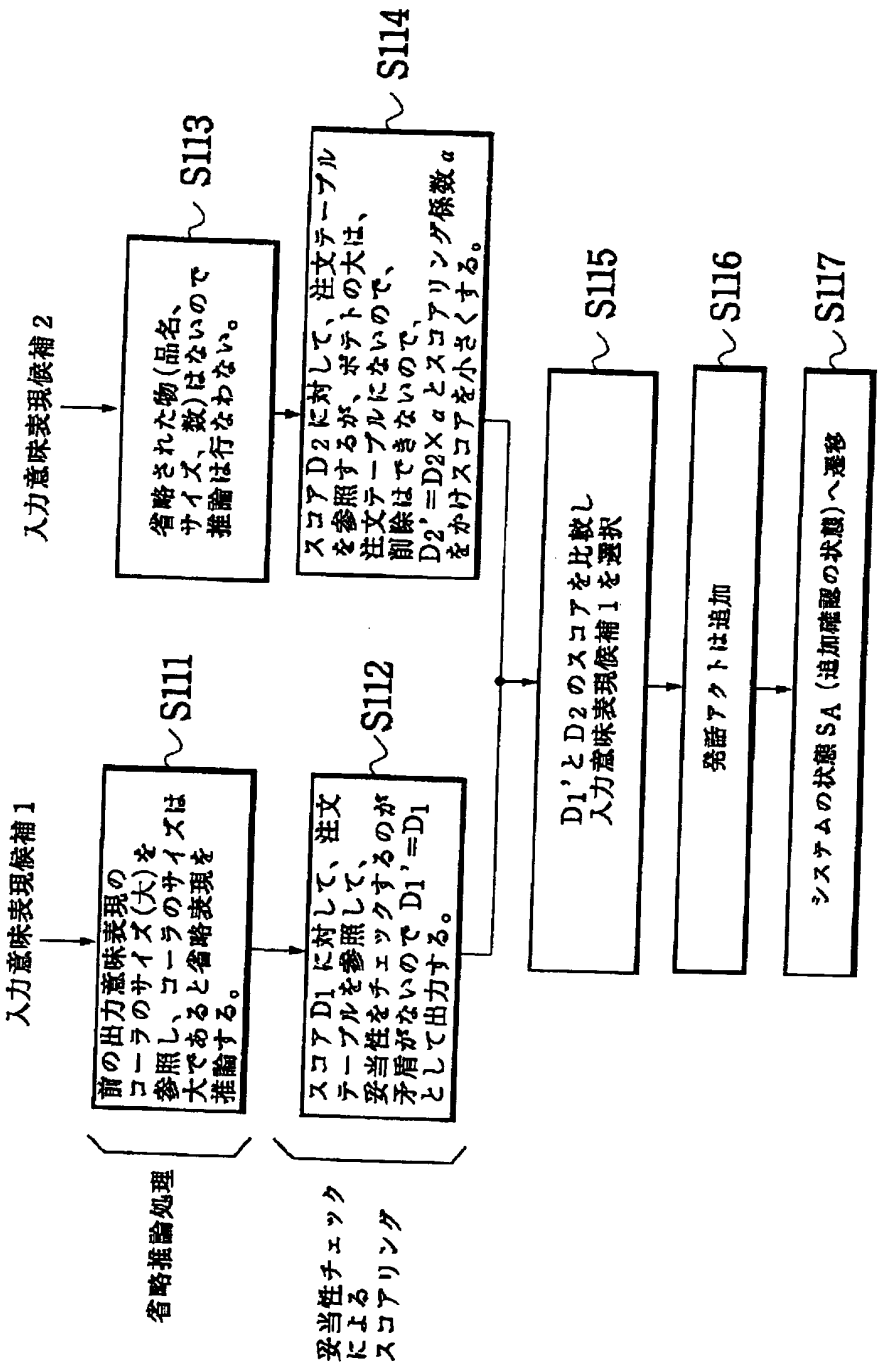
(d)入力意味表現候補１

ACT＝追加		D_1
品名	サイズ	個数
コーラ	?	1

(e)入力意味表現候補２

ACT＝削除		D_2
品名	サイズ	個数
ポテト	大	2

【図13】


入力意味表現候補1
前の出力意味表現のコーラのサイズ(大)を参照し、コーラのサイズは大であると省略表現を推論する。
S111
省略推論処理
スコアD1に対して、注文テーブルを参照して、妥当性をチェックするのが矛盾がないのでD1'=D1として出力する。
S112
妥当性チェックによるスコアリング
入力意味表現候補2
省略された物(品名、サイズ、数)はないので推論は行なわない。
S113
スコアD2に対して、注文テーブルを参照するが、ポテトの大は、注文テーブルにないので、前除はできないので、D2'=D2×αとスコアリング係数αをかけスコアを小さくする。
S114
D1'とD2のスコアを比較し入力意味表現候補1を選択
S115
発話アクトは追加
S116
システムの状態SA(追加確認の状態)へ遷移
S117

【図１６】

システムの応答の表現

全体確認：全注文の確認応答
例：ご注文は；
ハンバーガーを　１つ
コーヒーを　２つ
コーラの大を　４つ
ですね。

部分確認：全注文にかぎらず、幾つかの品目について確認応答
例：確認します。
ハンバーガーは　３つ
チーズバーガーは ２つ
ですね。

個別確認：１つづつ品目を確認応答
例：１つづつ確認します。
ハンバーガーは ２つ
ですね。

追加確認：追加確認応答
例：確認します。
コーヒーを ３つ
コーラの中を１つ
追加ですね。

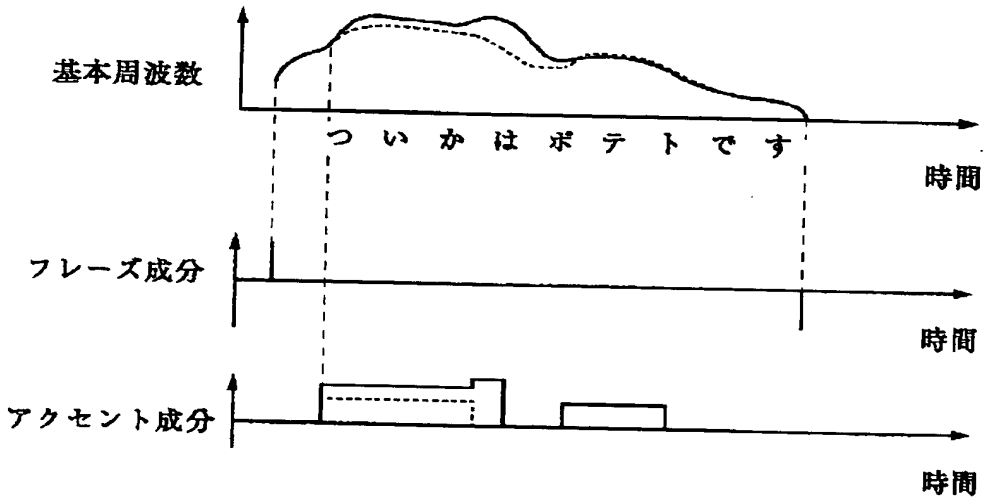
置換確認：置換の確認応答
例：確認します。
ハンバーガーが２つではなくチーズバーガーが ３つ
ですか。

再発話要求：初期化
例：すいません。
もう一度、最初からご注文をお願いします。

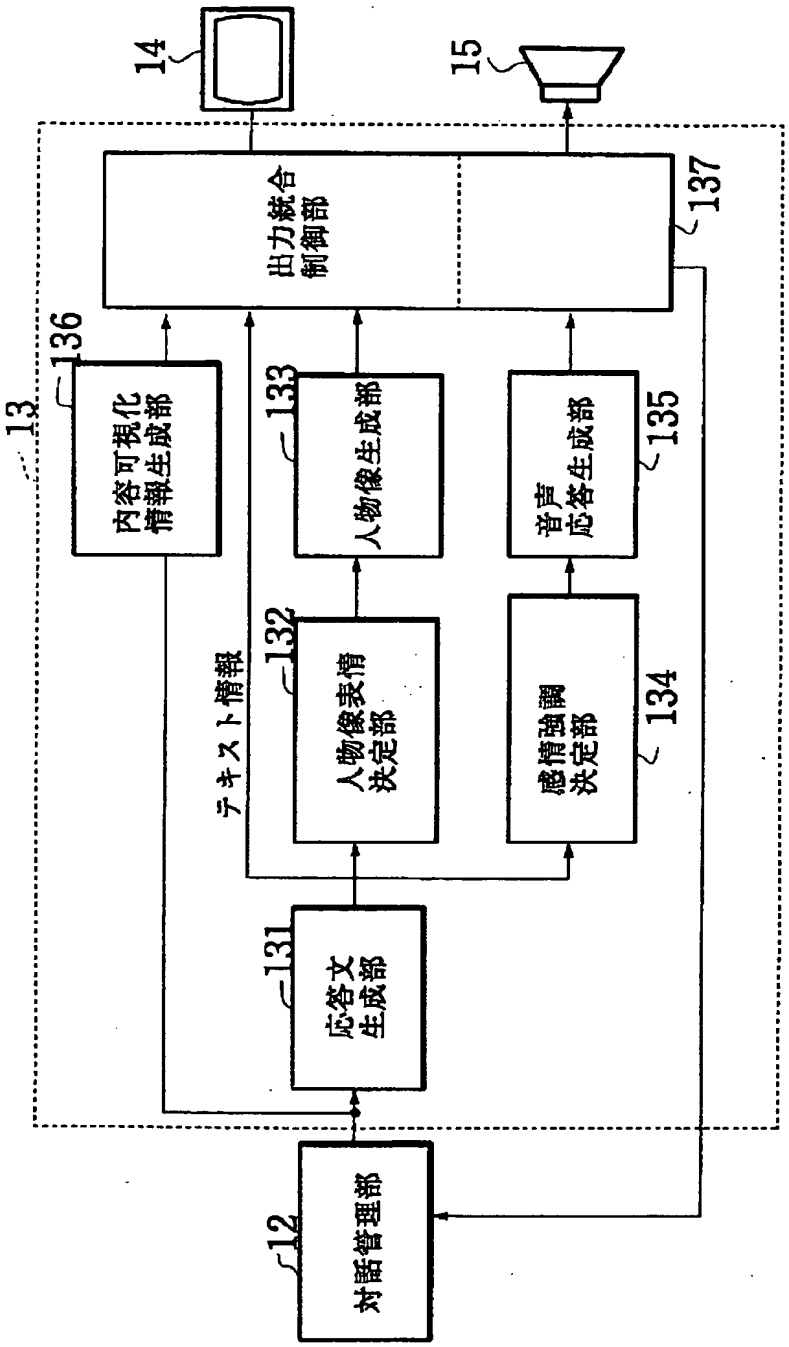
【図１８】

ACT＝部分確認		
品名	サイズ	個数
コーラ	大	1
ポテト	小	3
キーワードの尤度		
“コーラ” ＝0.8	“大”＝0.7	“1”＝0.4
“ポテト” ＝0.6	“小” ＝0.9	“3”＝0.8
ユーザの発声速度 ＝ 8 モーラ／秒		

【図２６】

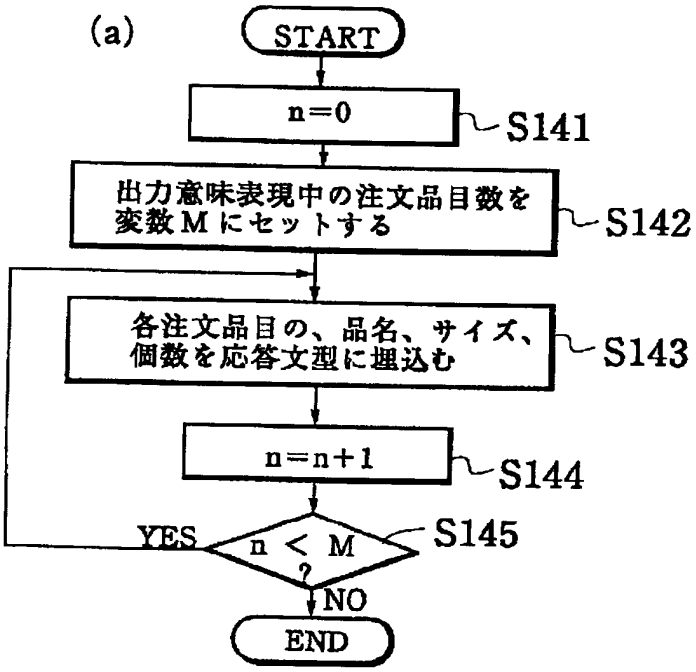


【図１９】


14
15
出力統合制御部
137
136
内容可視化情報生成部
13
133
人物像生成部
135
音声応答生成部
テキスト情報
132
人物像表情決定部
134
感情強調決定部
131
応答文生成部
12
対話管理部

【図２２】

(a)



(b)

ACT＝部分確認		
品名	サイズ	個数
コーラ	大	1
ポテト	小	3

(c)

ACT ＝ 部分確認

「確認します。〈品名〉〈サイズ〉〈個数〉 ですね。」

(d)

「確認します。 コーラの大を１つ、 ポテトの小を３つですね。」

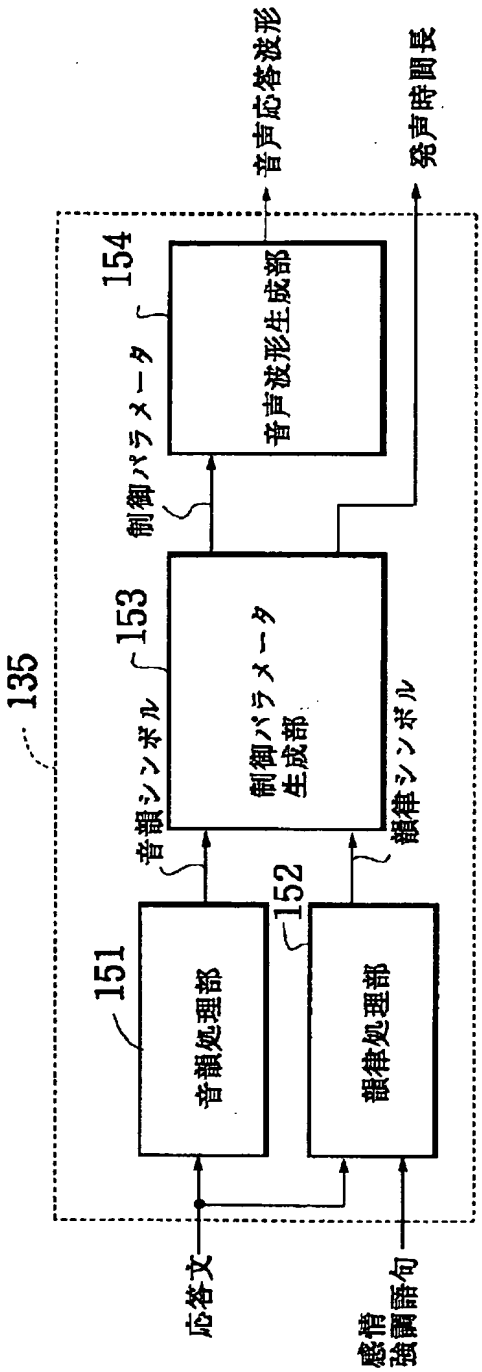
【図23】

システム状態	ユーザ状態	繰返し回数 N	確信度 D	人物像タイプ	表情
S0	U0	—	—	あいさつ	喜び
SP1	UP1	0≦N<2	0≦D<0.7	確認	戸惑い
SP1	UP1	0≦N<2	0.7≦D<1.0	確認	普通
SP1	UP1	2<N	0≦D<1.0	確認	申し分けない
.....	.....	.....	.....	.....	.....
S10	UP15	0≦N<2	0≦D<1.0	個別確認	普通
.....	.....	.....	.....	.....	.....
SP25	UP30	—	—	再発話要求	申し分けない
.....	.....	.....	.....	.....	.....

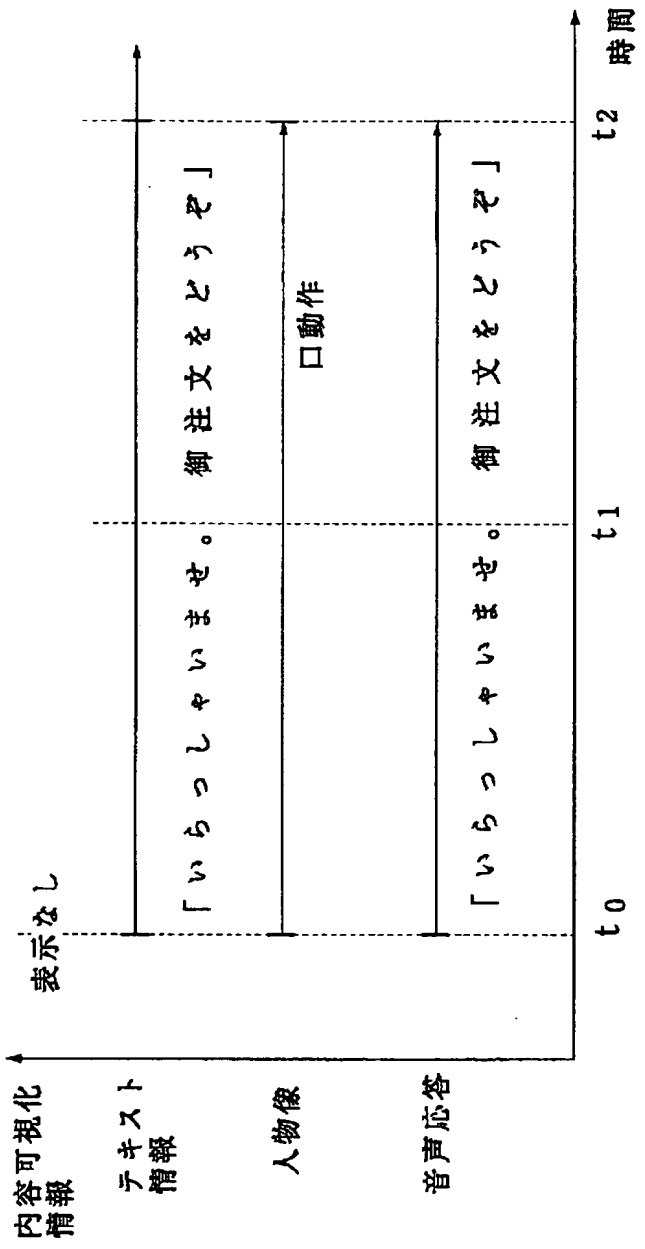
【図24】

システム状態	ユーザ状態	繰返し回数 N	確信度 D	人物像タイプ	感情表現
S0	U0	—	—	あいさつ	喜び
SP1	UP1	0≦N<2	0≦D<0.7	確認	戸惑い
SP1	UP1	0≦N<2	0.7≦D<1.0	確認	普通
SP1	UP1	2<N	0≦D<1.0	確認	申し分けない
.....	.....	.....	.....	.....	.....
S10	UP15	0≦N<2	0≦D<1.0	個別確認	普通
.....	.....	.....	.....	.....	.....
SP25	UP30	—	—	再発話要求	申し分けない
.....	.....	.....	.....	.....	.....

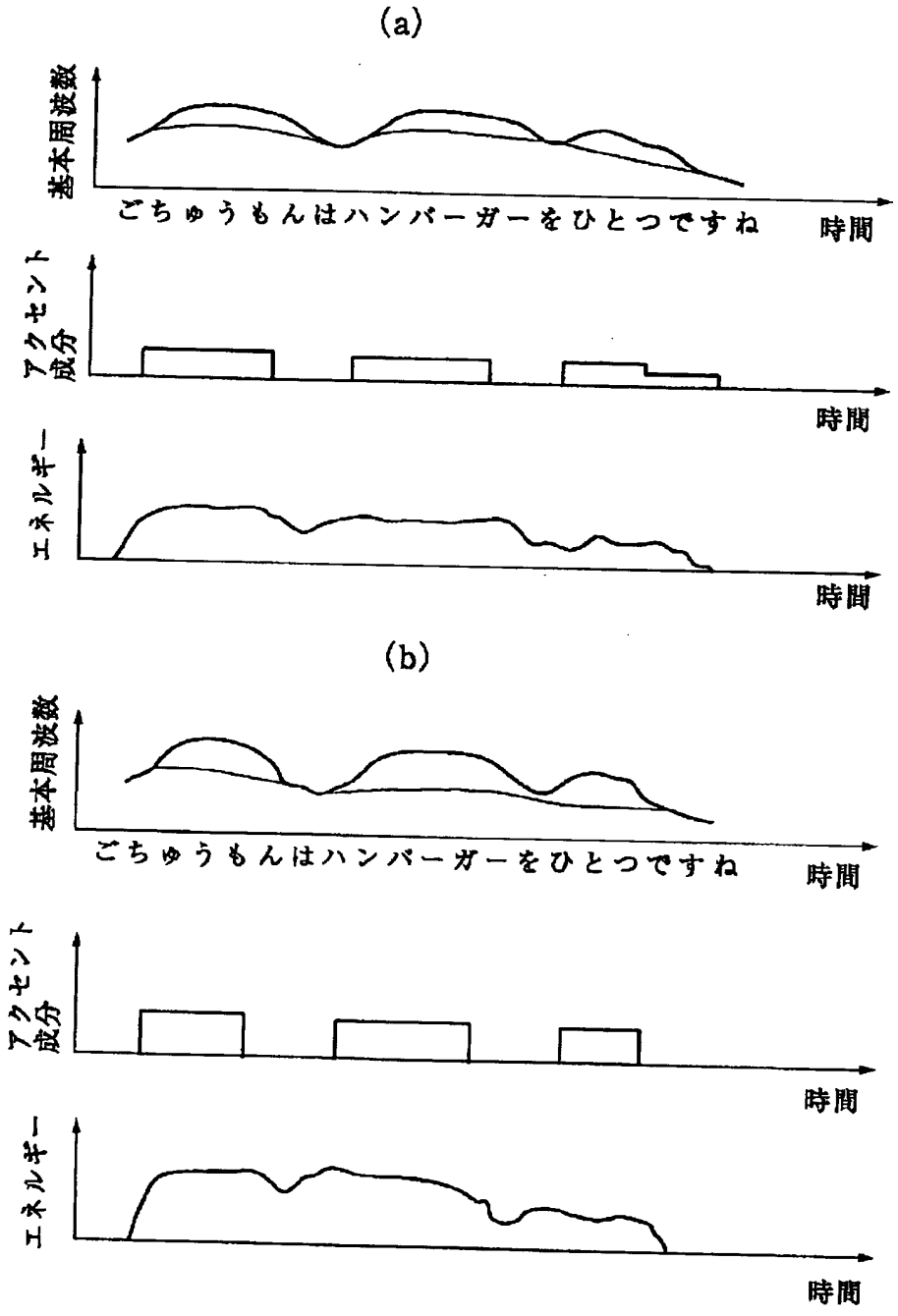
【図２５】


135
151
音韻処理部
152
韻律処理部
音韻シンボル
153
制御パラメータ
生成部
韻律シンボル
制御パラメータ
154
音声波形生成部
音声応答波形
発声時間長
応答文
感情
強調語句

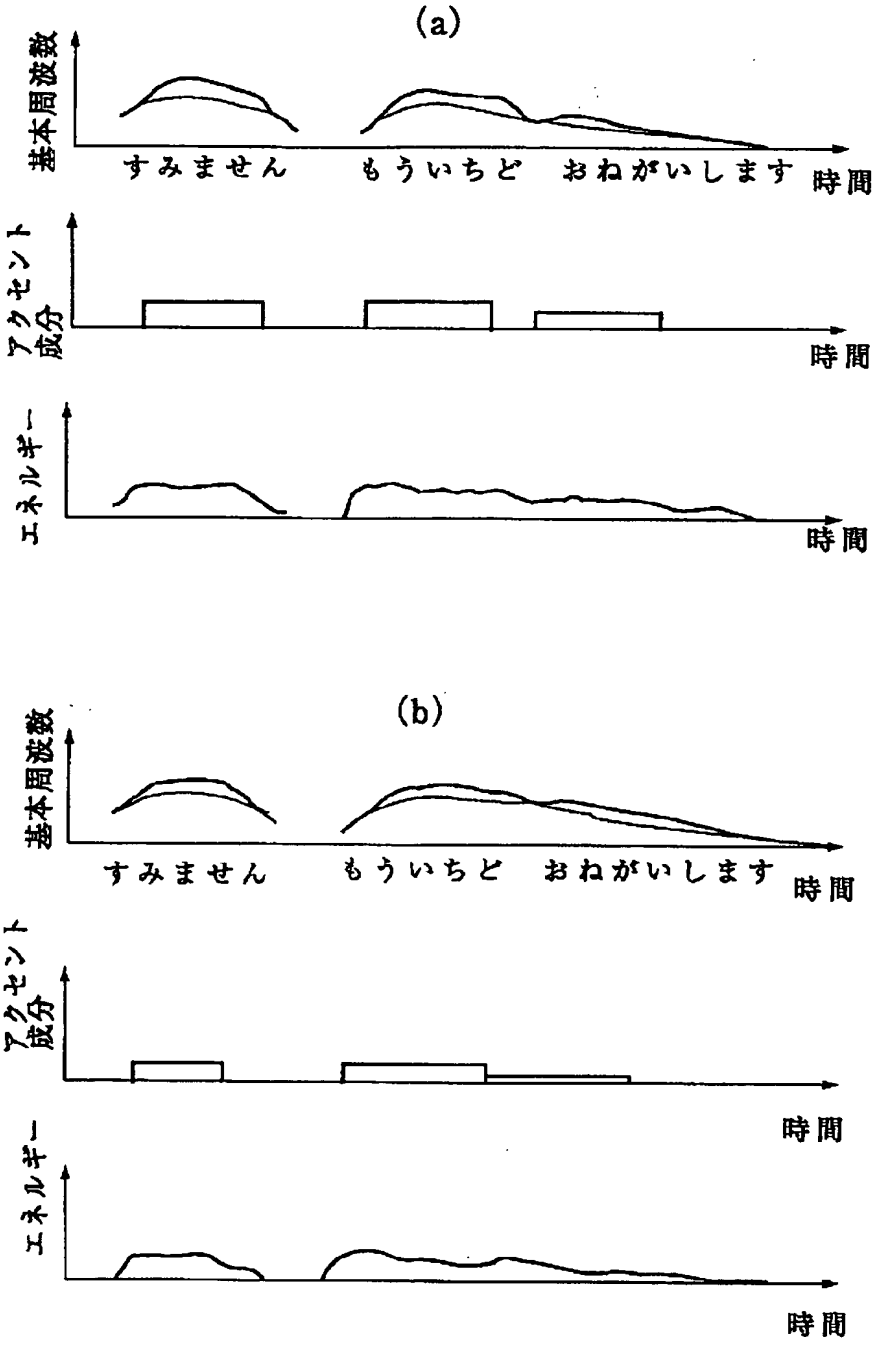
【図３０】


内容可視化
情報
テキスト
情報
人物像
音声応答
表示なし
「いらっしゃいませ。御注文をどうぞ」
口動作
「いらっしゃいませ。御注文をどうぞ」
t0
t1
t2
時間

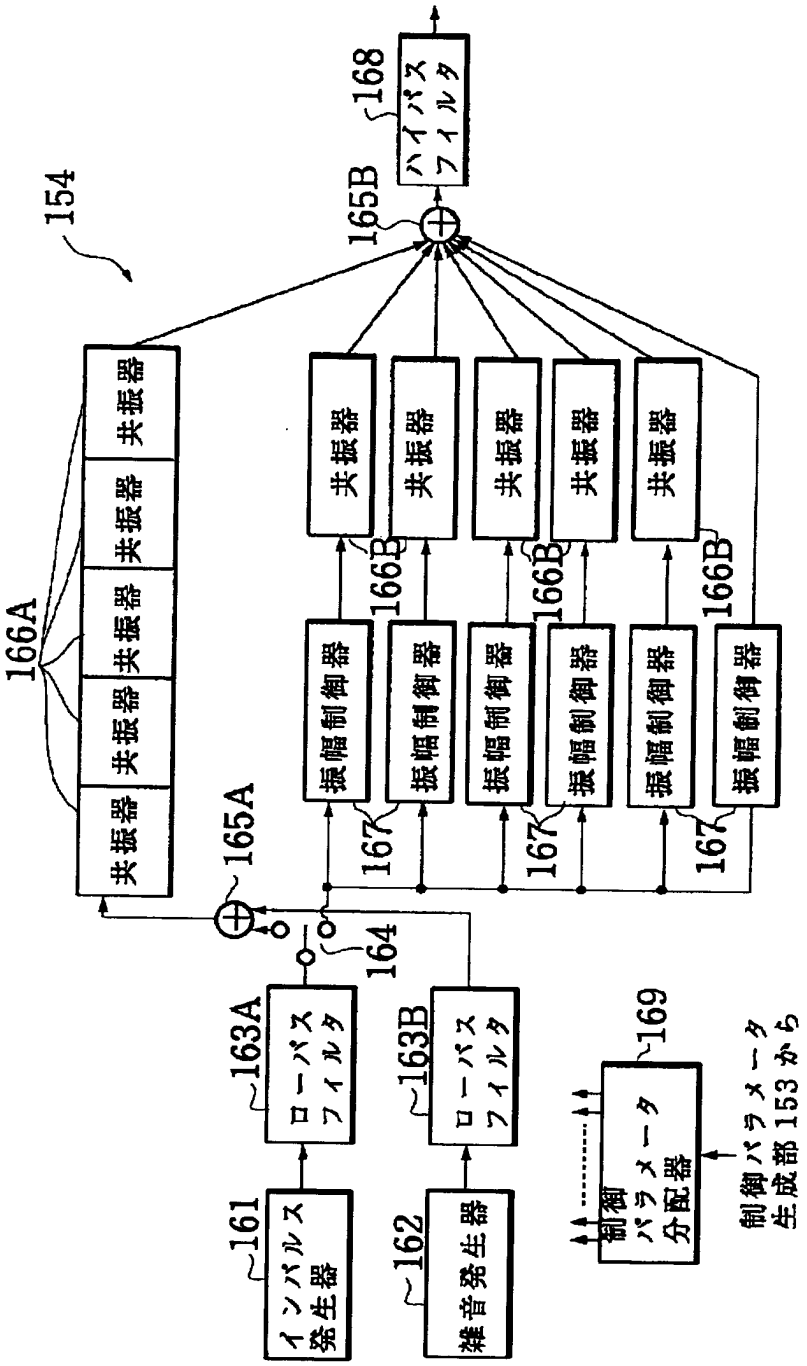
【図２７】


(a)
基本周波数
ごちゅうもんはハンバーガーをひとつですね
時間
アクセント成分
時間
エネルギー
時間
(b)
基本周波数
ごちゅうもんはハンバーガーをひとつですね
時間
アクセント成分
時間
エネルギー
時間

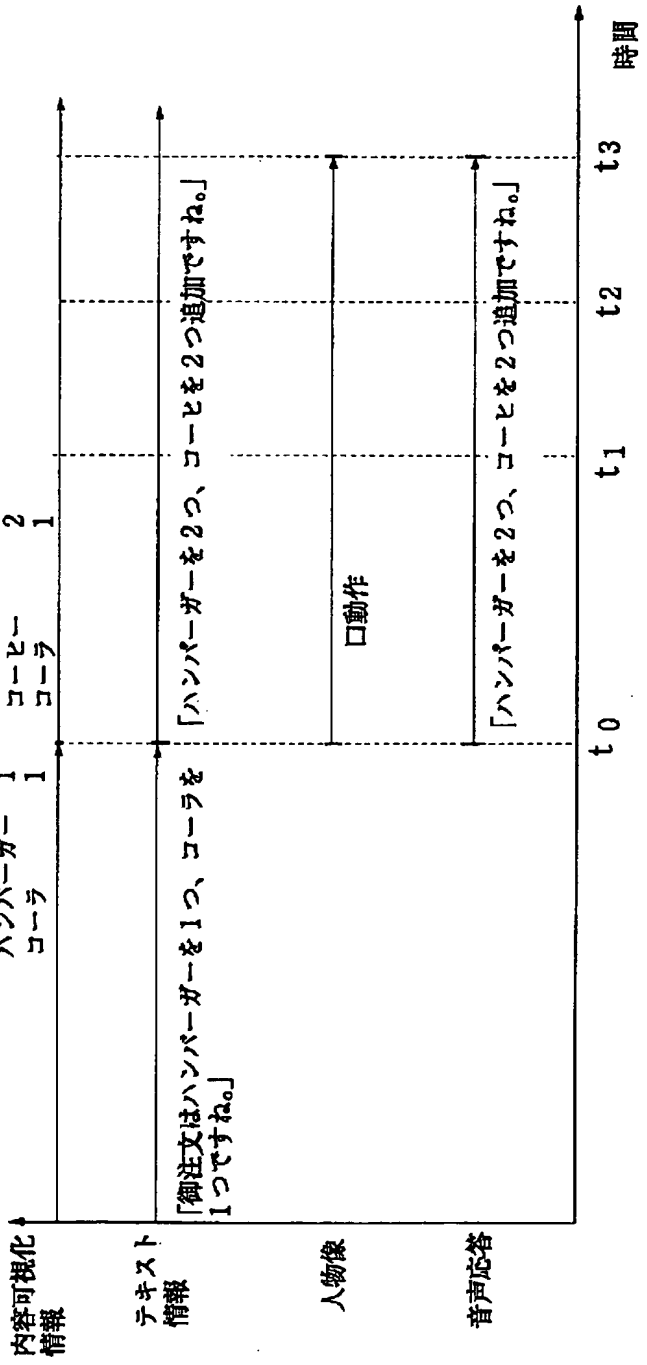
【図２８】


(a)
基本周波数
すみません
もういちど
おねがいします
時間
アクセント成分
時間
エネルギー
時間
(b)
基本周波数
すみません
もういちど
おねがいします
時間
アクセント成分
時間
エネルギー
時間

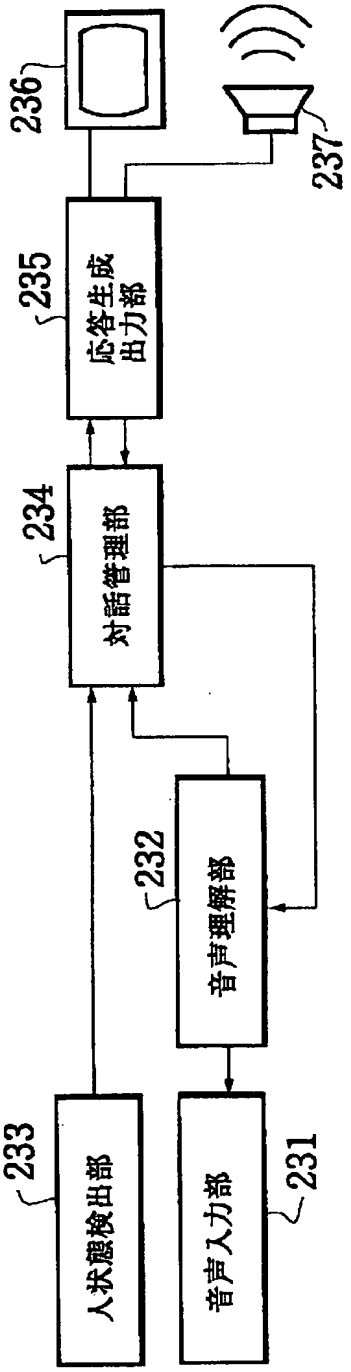
【図２９】


154
166A
共振器
共振器
共振器
共振器
共振器
165A
164
163A
ローパスフィルタ
163B
ローパスフィルタ
161
インパルス発生器
162
雑音発生器
169
制御パラメータ分配器
制御パラメータ生成部153から
166B
振幅制御器
167
165B
168
ハイパスフィルタ

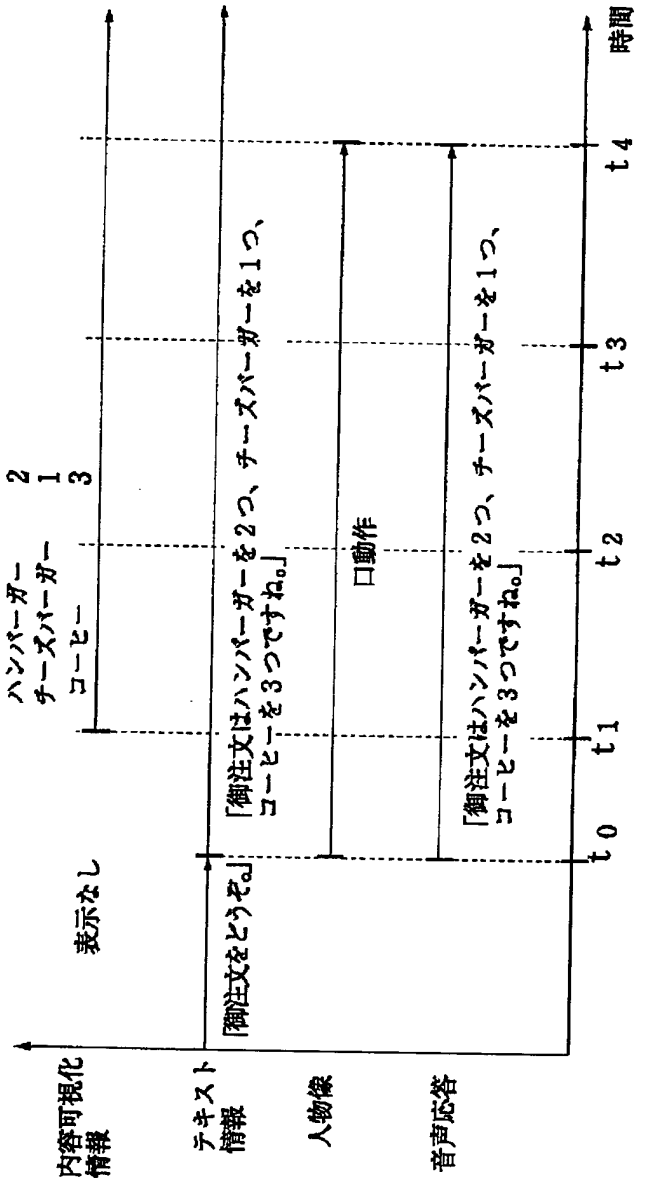
【図３１】


内容可視化情報
テキスト情報
人物像
音声応答
時間
t0
t1
t2
t3
ハンバーガー　1
コーラ　1
ハンバーガー　3
コーヒー　2
コーラ　1
「御注文はハンバーガーを1つ、コーラを1つですね。」
「ハンバーガーを2つ、コーヒを2つ追加ですね。」
口動作
「ハンバーガーを2つ、コーヒを2つ追加ですね。」

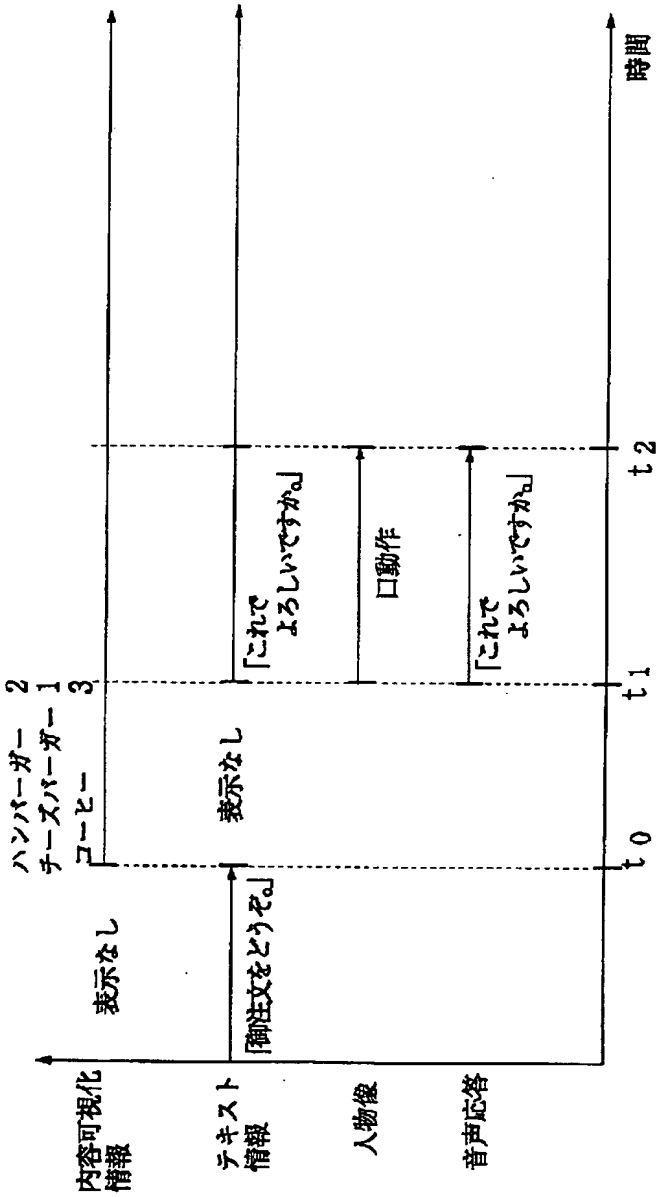
【図４２】


236
237
235
応答生成出力部
234
対話管理部
232
音声理解部
233
人状態検出部
音声入力部
231

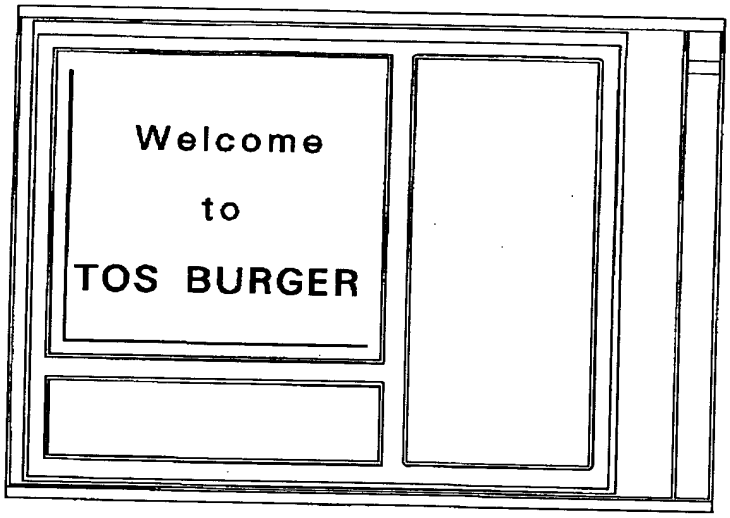
【図３２】


内容可視化情報
テキスト情報
人物像
音声応答
表示なし
ハンバーガー　2
チーズバーガー　1
コーヒー　3
「御注文をどうぞ。」
「御注文はハンバーガーを２つ、チーズバーガーを１つ、コーヒーを３つですね。」
口動作
「御注文はハンバーガーを２つ、チーズバーガーを１つ、コーヒーを３つですね。」
t0
t1
t2
t3
t4
時間

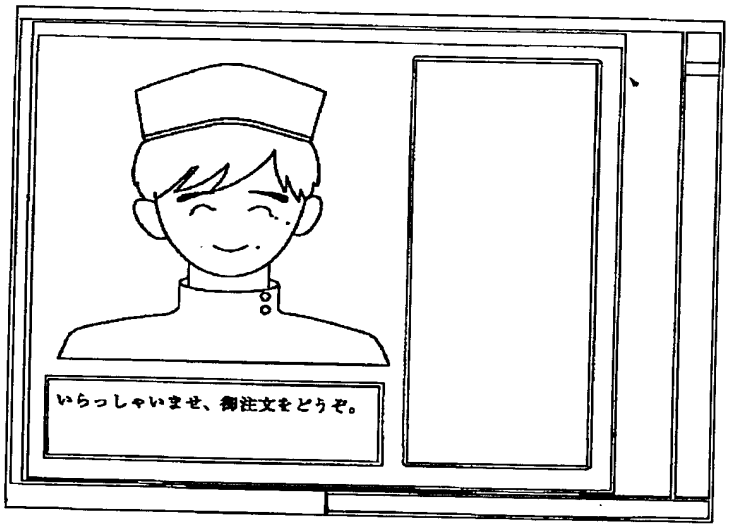
【図３３】


内容可視化情報
テキスト情報
人物像
音声応答
時間
t0
t1
t2
表示なし
「御注文をどうぞ」
ハンバーガー 2
チーズバーガー 1
コーヒー 3
表示なし
「これでよろしいですか」
口動作
「これでよろしいですか」

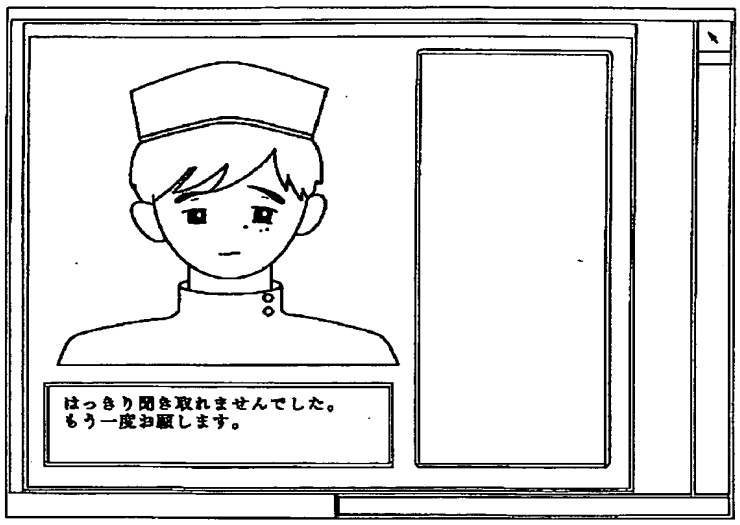
【図３４】



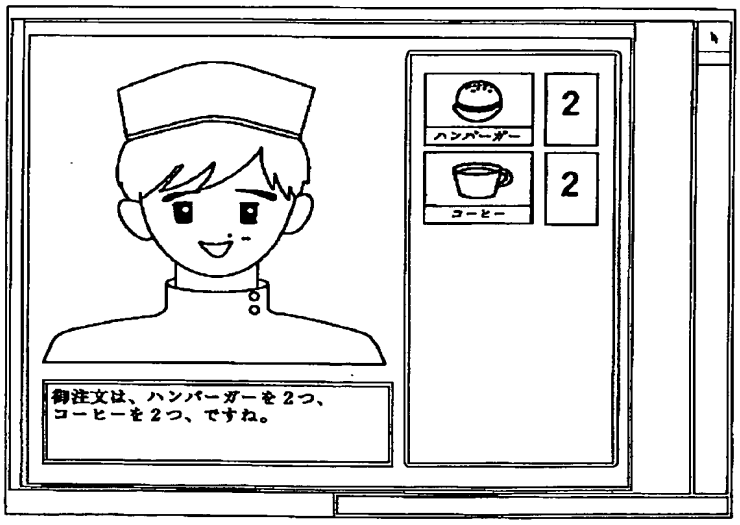
【図３５】



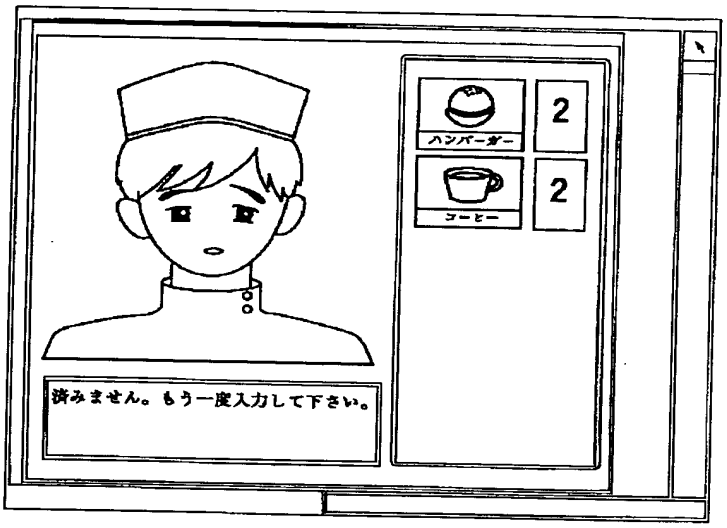
【図３６】


はっきり聞き取れませんでした。
もう一度お願します。

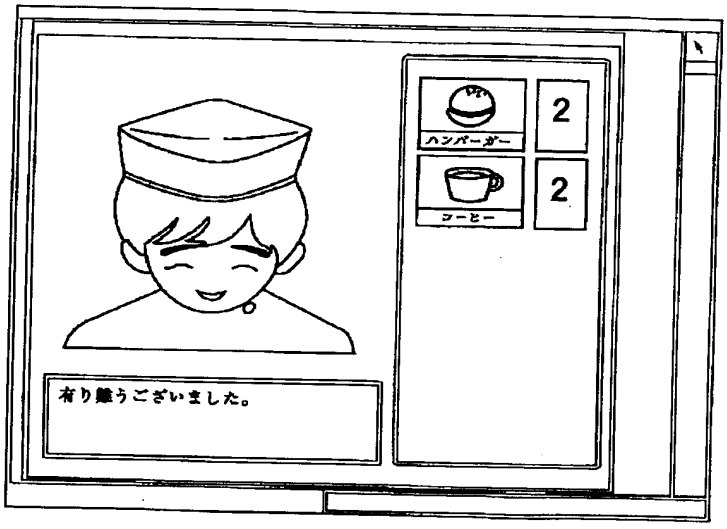
【図３７】


ハンバーガー
2
コーヒー
2
御注文は、ハンバーガーを２つ、
コーヒーを２つ、ですね。

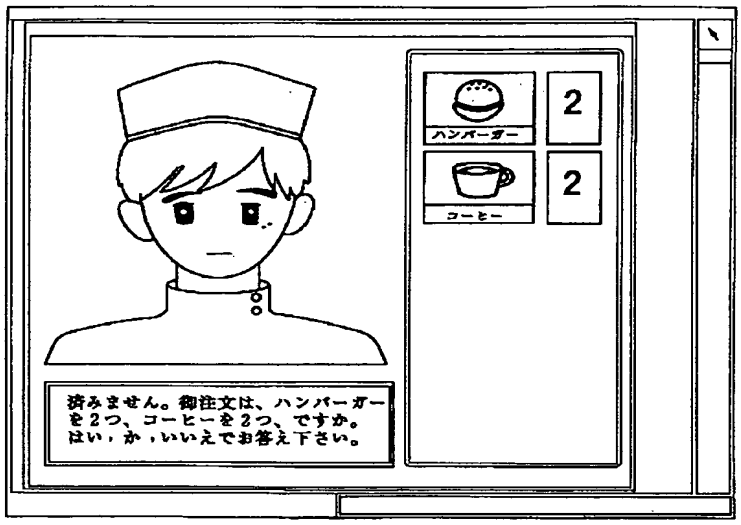
【図３８】


ハンバーガー
2
コーヒー
2
済みません。もう一度入力して下さい。

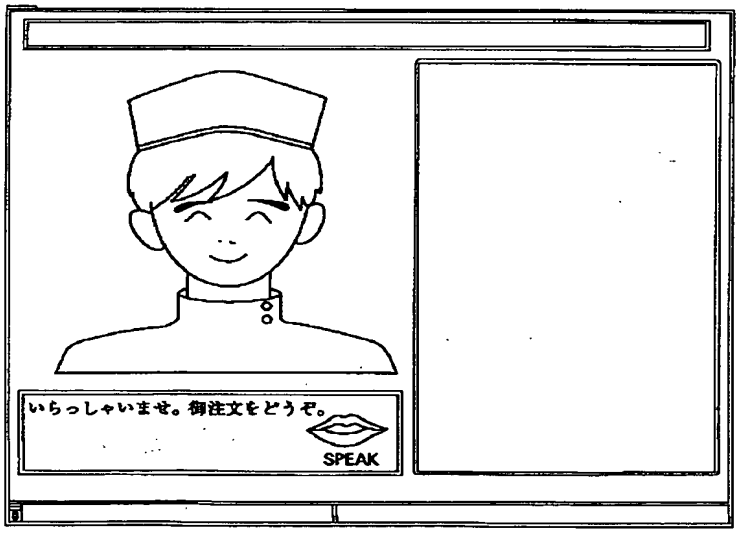
【図３９】


ハンバーガー
2
コーヒー
2
有り難うございました。

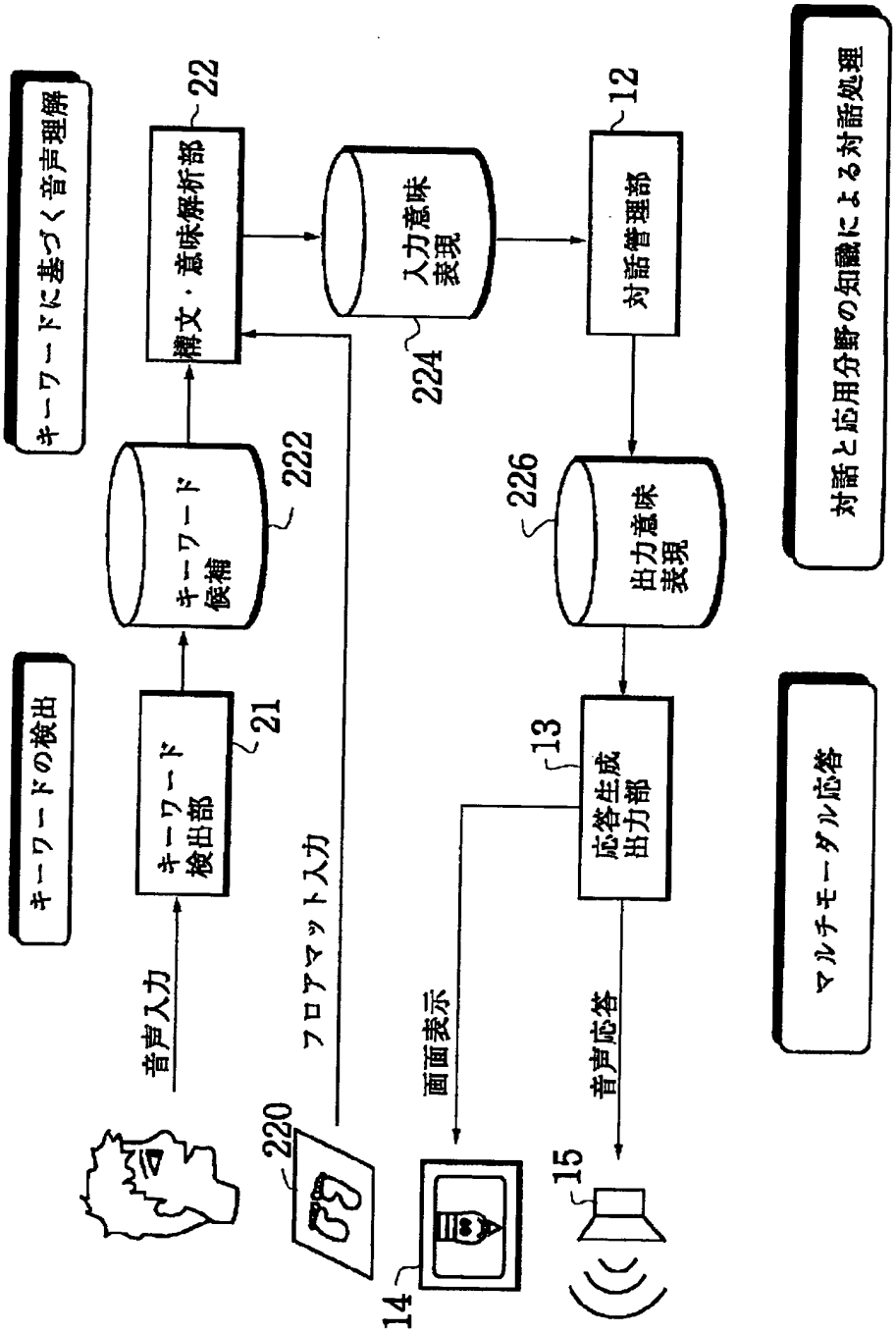
【図４０】


ハンバーガー
2
コーヒー
2
済みません。御注文は、ハンバーガーを2つ、コーヒーを2つ、ですか。はい，か，いいえでお答え下さい。

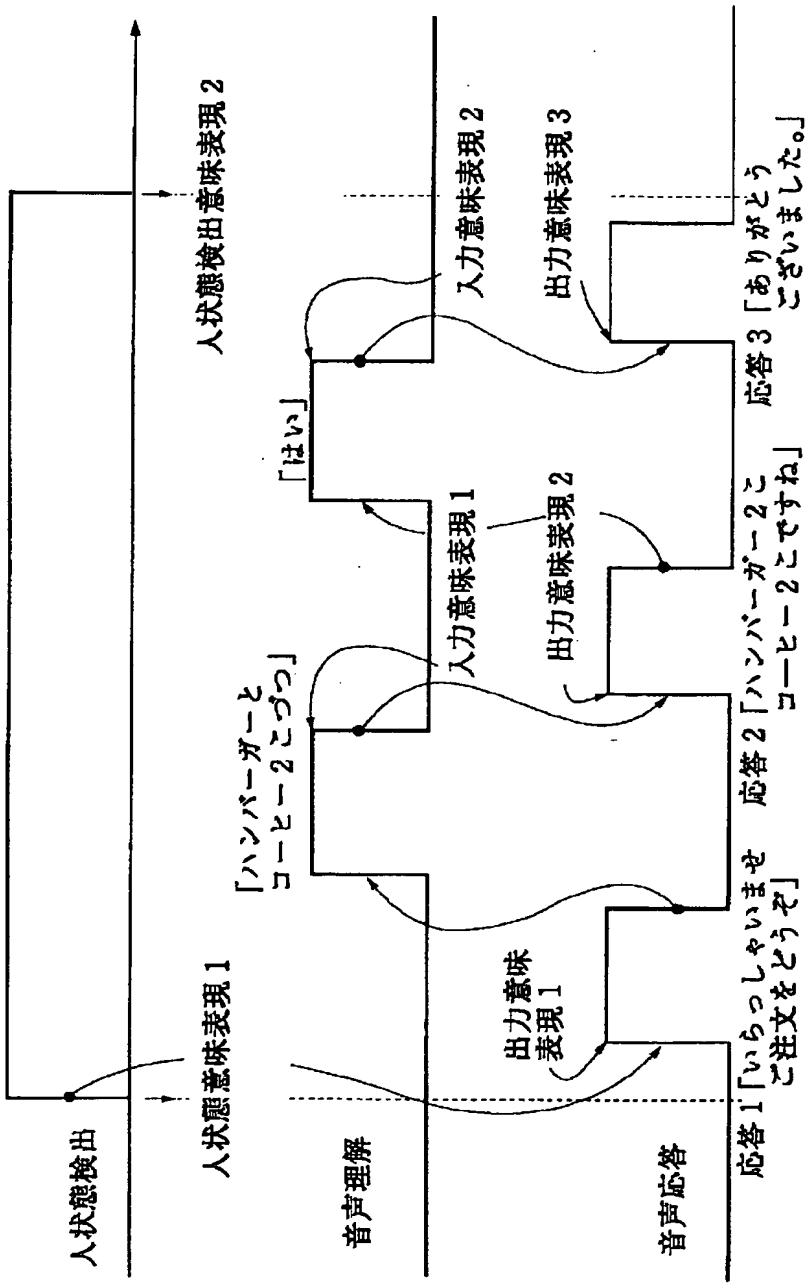
【図４９】


いらっしゃいませ。御注文をどうぞ。
SPEAK

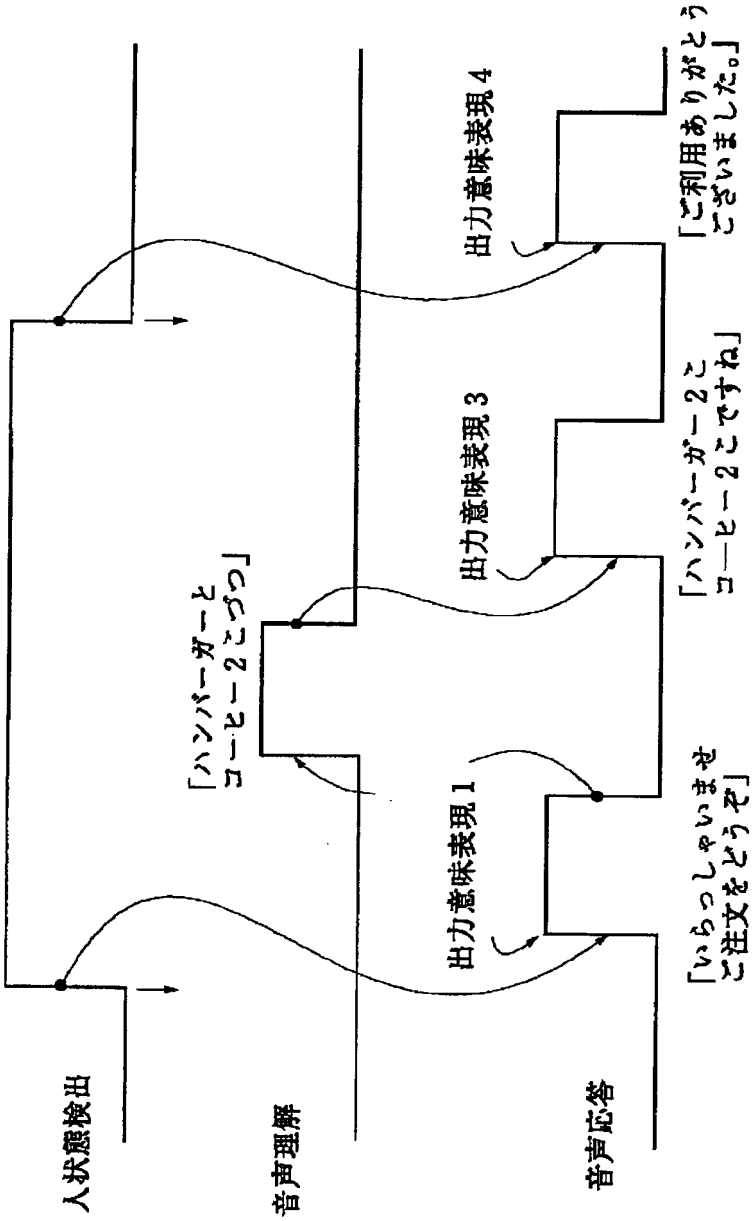
【図41】


キーワードに基づく音声理解
キーワードの検出
マルチモーダル応答
対話と応用分野の知識による対話処理
22
12
構文・意味解析部
入力意味表現
対話管理部
224
キーワード候補
222
226
出力意味表現
キーワード検出部
21
13
応答生成出力部
音声入力
フロアマット入力
画面表示
音声応答
220
14
15

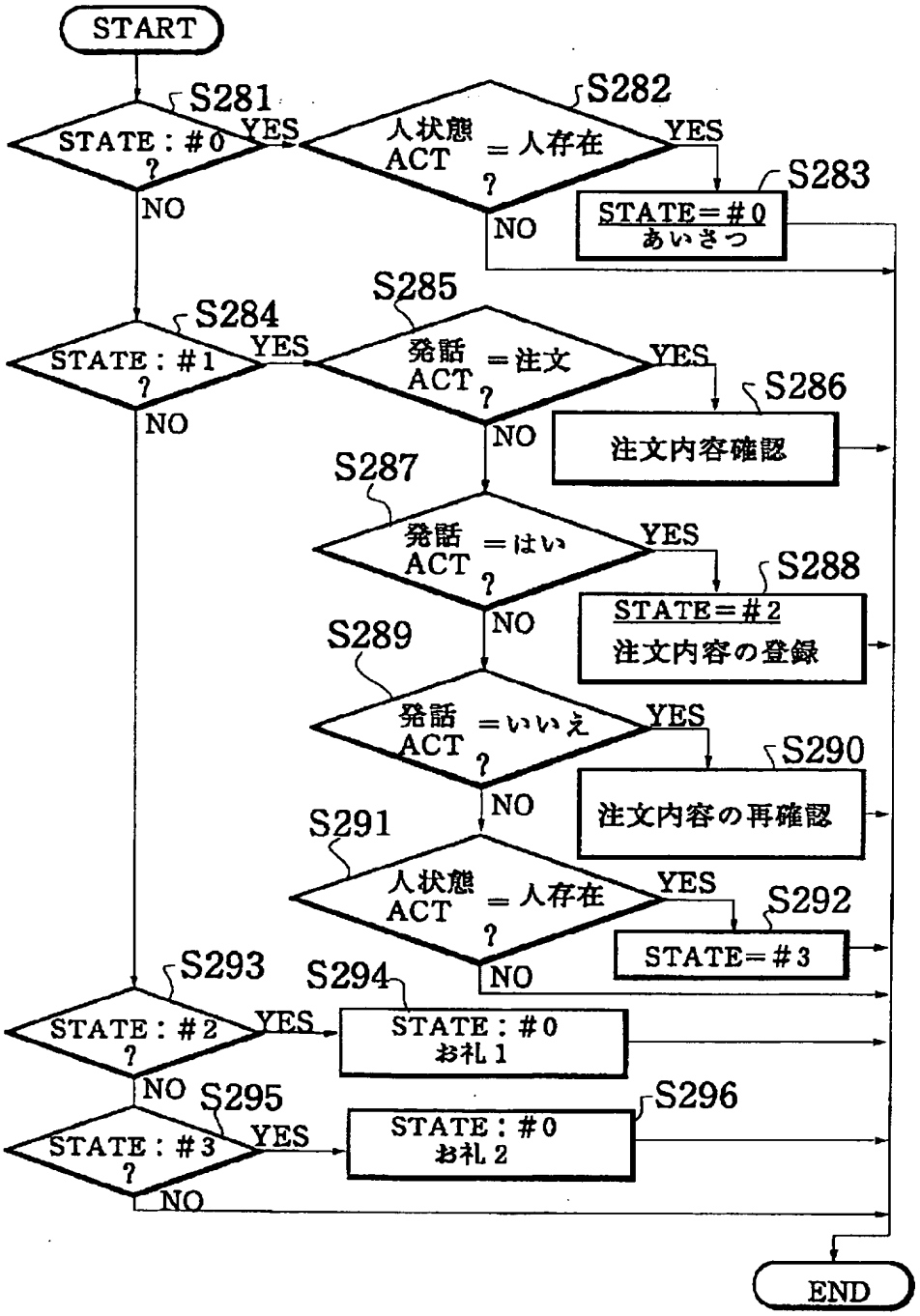
【図４４】


人状態検出
人状態意味表現１
人状態検出意味表現２
音声理解
「ハンバーガーとコーヒー２こづつ」
「はい」
入力意味表現１
入力意味表現２
出力意味表現１
出力意味表現２
出力意味表現３
音声応答
応答１「いらっしゃいませ ご注文をどうぞ」
応答２「ハンバーガー２こ コーヒー２こですね」
応答３「ありがとうございました。」

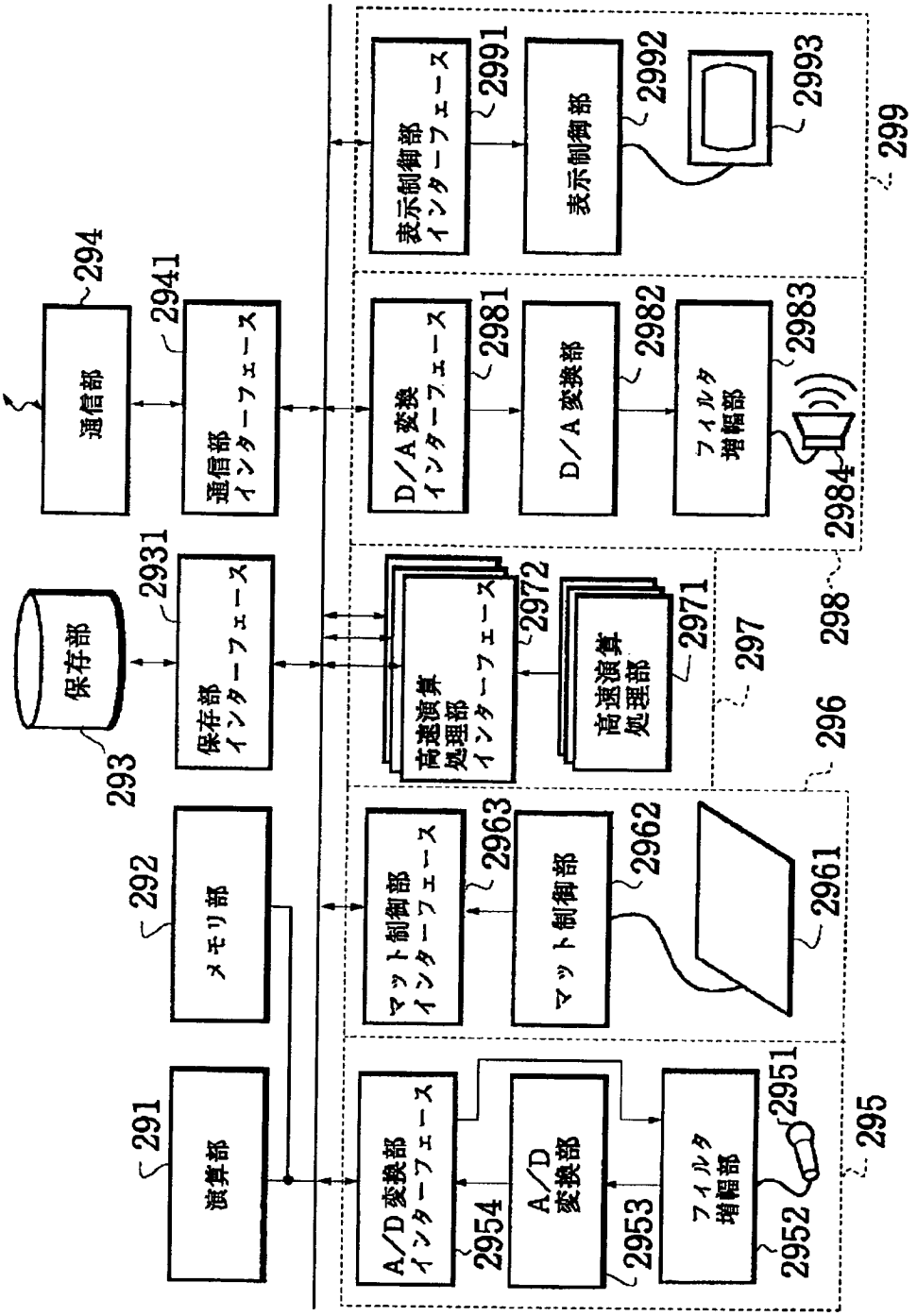
【図４５】


人状態検出
音声理解
音声応答
「ハンバーガーと
コーヒー２こづつ」
出力意味表現１
出力意味表現３
出力意味表現４
「いらっしゃいませ
ご注文をどうぞ」
「ハンバーガー２こ
コーヒー２こですね」
「ご利用ありがとう
ございました。」

【図４６】

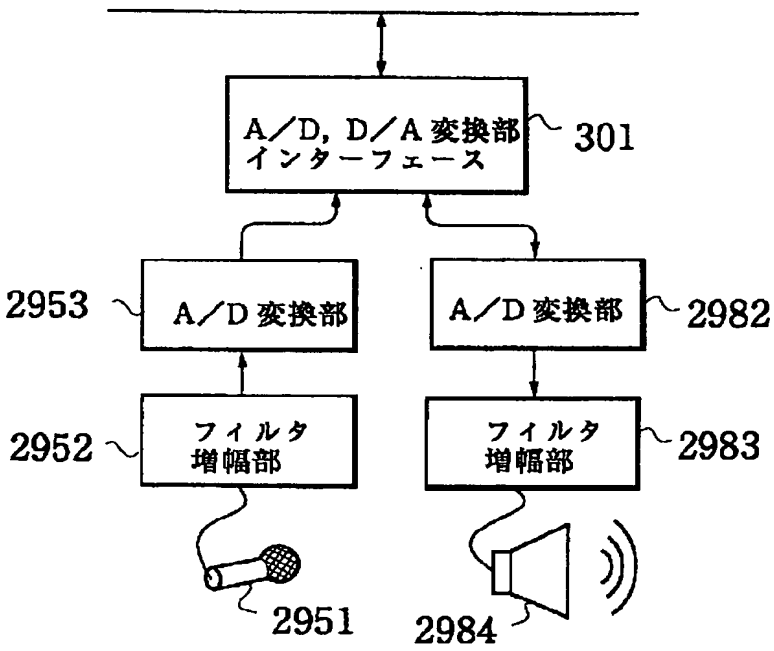

START
S281
STATE：#0 ?
YES
NO
S282
人状態 ACT ＝人存在 ?
YES
NO
S283
STATE＝#0
あいさつ
S284
STATE：#1 ?
YES
NO
S285
発話 ACT ＝注文 ?
YES
NO
S286
注文内容確認
S287
発話 ACT ＝はい ?
YES
NO
S288
STATE＝#2
注文内容の登録
S289
発話 ACT ＝いいえ ?
YES
NO
S290
注文内容の再確認
S291
人状態 ACT ＝人存在 ?
YES
NO
S292
STATE＝#3
S293
STATE：#2 ?
YES
NO
S294
STATE：#0
お礼１
S295
STATE：#3 ?
YES
NO
S296
STATE：#0
お礼２
END

【図４７】

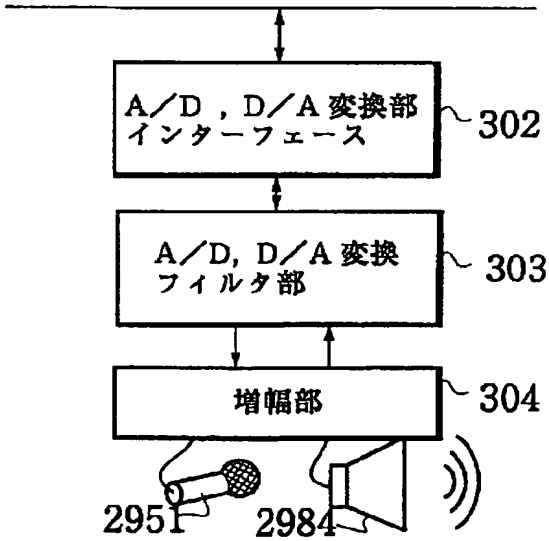

299
表示制御部インターフェース
2991
表示制御部
2992
2993
298
D／A変換インターフェース
2981
D／A変換部
2982
フィルタ増幅部
2983
2984
297
高速演算処理部インターフェース
2972
高速演算処理部
2971
296
マット制御部インターフェース
2963
マット制御部
2962
2961
295
A／D変換部インターフェース
2954
A／D変換部
2953
フィルタ増幅部
2951
2952
通信部
294
通信部インターフェース
2941
保存部
293
保存部インターフェース
2931
メモリ部
292
演算部
291

【図４８】

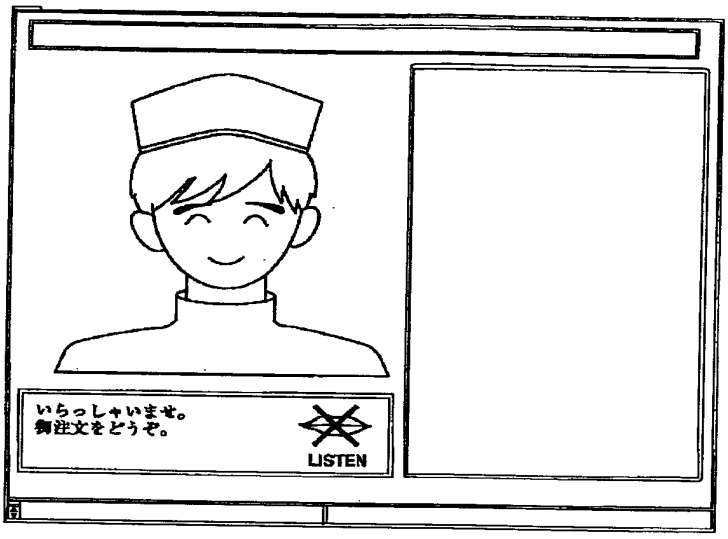
(a)


A／D，D／A 変換部
インターフェース
301
2953
A／D 変換部
A／D 変換部
2982
2952
フィルタ
増幅部
フィルタ
増幅部
2983
2951
2984

(b)


A／D，D／A 変換部
インターフェース
302
A／D，D／A 変換
フィルタ部
303
増幅部
304
2951
2984

【図５０】



フロントページの続き

(51)Int.Cl.⁵	識別記号	庁内整理番号	F I	技術表示箇所
Ｇ１０Ｌ　3/00		５７１　Ｈ　8842－5H		

(72)発明者　坪井　宏之
神奈川県川崎市幸区小向東芝町１　株式会社東芝研究開発センター内
(72)発明者　貞本　洋一
神奈川県川崎市幸区小向東芝町１　株式会社東芝研究開発センター内
(72)発明者　山下　泰樹
神奈川県川崎市幸区小向東芝町１　株式会社東芝研究開発センター内
(72)発明者　永田　仁史
神奈川県川崎市幸区小向東芝町１　株式会社東芝研究開発センター内
(72)発明者　瀬戸　重宣
神奈川県川崎市幸区小向東芝町１　株式会社東芝研究開発センター内
(72)発明者　新地　秀昭
東京都青梅市新町1385番地　東芝ソフトウェアエンジニアリング株式会社内
(72)発明者　橋本　秀樹
東京都青梅市新町1385番地　東芝ソフトウェアエンジニアリング株式会社内